The transcriptome of barley chloroplasts revealed by deep sequencing

DISSERTATION

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m

(Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

M. Sc. Petya Zhelyazkova

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Andreas Herrmann

Gutachter:
1. Prof. Dr. Thomas Börner
2. Prof. Dr. Ralf Bock
3. Prof. Dr. Wolfgang Schuster

Tag der mündlichen Prüfung: 28.09.2012

ABSTRACT

The current view on plastid gene expression is mainly based on the analysis of a few individual genes, and thus it is lacking in comprehensiveness. Here, a novel differential RNA-seq approach, designed to discriminate between primary and processed transcripts, was used to obtain a deeper insight into the plastid transcription and RNA maturation of mature barley (*Hordeum vulgare* L.) chloroplasts.

Transcription in plastids of higher plants is dependent on two different transcription machineries, a plastid-encoded bacterial-type RNA polymerase (PEP) and a nuclear-encoded phage-type RNA polymerase (NEP), which recognize distinct types of promoters. This study provided a thorough investigation into the distribution of transcription start sites within the plastid genome of green (mature chloroplasts; transcription by both PEP and NEP) and white (PEP-deficient plastids; transcription by NEP) plastids of the barley line albostrians. This analysis led to new insights on polymerase specific gene expression in plastids.

Recent studies have suggested that non-coding RNAs (ncRNAs) are common in chloroplasts. However, they did not directly detect ncRNAs generated via transcription, the so far most abundant class of known regulatory ncRNAs in bacteria. Here, dRNA-seq analysis of the transcriptome of barley chloroplasts demonstrated the existence of numerous ncRNA generated via transcription of free-standing genes.

Major events in plastid mRNA maturation include 5' and 3' processed end formation and intercistronic processing. Recently, a PPR (pentatricopeptide repeat) protein was shown to participate in the generation of several plastid mRNA processed ends by serving as a barrier to exonucleases. This study provided evidence for the global impact of this mechanism on processed termini formation in chloroplasts.

Keywords: plastids, transcription, plastid-encoded RNA polymerase, nuclear-encoded RNA polymerase, non-coding RNAs, mRNA maturation.

ZUSAMMENFASSUNG

Die gegenwärtige Vorstellung von Genexpression in Plastiden leitet sich von der Analyse weniger, individueller Gene ab und ist deshalb noch relativ lückenhaft. In dieser Arbeit sollte daher differenzierende RNA Sequenzierung- eine neue Methode, die zwischen prozessierten und Primärtranskripten unterscheiden kann, verwendet werden, um ein vollständigeres Bild des Transkriptionsprozesses und der RNA Prozessierung von Hordeum vulgare L. (Gerste) Chloroplasten zu erhalten.

Plastidengene in höheren Pflanzen können sowohl von einer plastidenkodierten, bakterienähnlichen RNA-Polymerase (PEP), als auch von einer kernkodierten, phagenähnlichen RNA-Polymerase (NEP), die beide unterschiedliche Promotoren erkennen, abgelesen werden. In dieser Arbeit wurde die Verteilung von Transkriptionsstartstellen innerhalb des Plastidengenoms von grünen (reife Chloroplasten; Transkriptionsaktivität von PEP und NEP) und weißen Plastiden (Transkriptionsaktivität von NEP) der Gerstenmutantenlinie albostrians analysiert. Dies führte zu neuen Erkenntnissen bezüglich polymerasenspezifischer Genexpression in Plastiden.

Auf Grundlage neuerer Arbeiten wird angenommen, daß nicht kodierende RNAs (ncRNAs) in Chloroplasten vorkommen. Die bisher verwendeten Methoden waren jedoch nicht geeignet, ncRNAs als Primärtranskripte zu identifizieren, die zumindest in Prokaryoten die häufigste Klasse von ncRNAs darstellen. In dieser Arbeit konnte durch dRNA-seq gezeigt werden, daß auch in Plastiden zahlreiche ncRNAs als Primärtranskripte generiert werden.

Die wichtigsten Schritte im Prozess der mRNA Reifung in Plastiden sind 5´und 3´ Endformation und intercistronische Prozessierung. Vor Kurzem wurde gezeigt, daß ein PPR (Pentatricopeptide repeat) Protein zur Bildung der Ende von einigen prozessierten Plastiden mRNAs beiträgt, indem es als Hindernis für Exonukleasen wirkt. Mit dieser Arbeit konnte gezeigt werden, daß dies ein genereller Mechanismus zur Bildung prozessierter mRNA-Enden in Chloroplasten ist.

Schlagwörter: Plastiden, Transkription, plastidenkodierte RNA-Polymerase, kernkodierte RNA-Polymerase; nicht kodierende RNAs; mRNA Reifung.

TABLE OF CONTENTS

1 INTRODUCTION

Plants and algae are photoautotrophs, and thus the most important primary producers in our biosphere. They use the energy of sunlight, and water as an electron donor to fix carbon and release oxygen through a process known as photosynthesis. Therefore, plants and algae form the basis for virtually all the world's food and fuel, and together with cyanobacteria, photosynthetic bacteria, they supply the oxygen in our atmosphere. Chloroplasts, unique organelles of plants and algae, are the sites of photosynthesis, and as such most life on Earth depends on them.

1.1 The structure and function of plastids

Plastids are the characteristic organelles of land plants and algae. They can be grouped into several plastid types, which play different roles in the production and storage of important compounds for the plant cell. Chloroplasts are the most prominent form of plastids and constitute the basis of the autotrophic lifestyle due to their photosynthetic activity. Chloroplasts are surrounded by a double envelope membrane, which serves not only in transport (export/import) of metabolites but also participates in biochemical synthesis and coordination of plastid and nuclear gene expression). The stroma is the cytosol of plastids. The light dependent reactions of photosynthesis is mediated by four large protein complexes, photosystem I (PSI), photosystem II (PSII), cytochrome b_6/f and ATPase, embedded in the thylakoid membrane of chloroplasts. The photosynthetic thylakoid membrane is an extensively folded and structurally heterogeneous membrane that surrounds the thylakoid lumen, an aqueous environment that plays a vital role in photophosphorylation. Two main thylakoid structures are present: the grana - stacks of thylakoid discs, and the lamellae - thylakoids interconnecting the grana. PSII is mostly found in the grana membranes, PSI and ATPase mostly in the lamellae, while cytochrome b_6/f complex is distributed evenly between the two thylakoid structures (Dekker and Boekema, 2005). Apart from photosynthesis, chloroplasts also play essential roles in the synthesis of fatty acids, chlorophyll and other tetrapyrroles, starch, etc. (Neuhaus and Emes, 2000).

Non-photosynthetic plastids are also central players in the plant cell metabolism. The necessity of plants to attract insects and mammals, essential for reproduction and seed dispersal, let to the development of a brightly colored class of plastids, the chromoplasts, which synthesize and accumulate pigments (Lopez-Juez and Pyke, 2005). Leucoplasts, a group of plastids with no

pigments, are very widely distributed and function mainly as storage compartments of different molecules. For example, amyloplasts are a form of leucoplast specialized in the synthesis and accumulation of starch (Pyke, 2007).

1.2 The origin and evolution of plastids

1.2.1 History of the endosymbiotic theory

Since more than hundred years, scientists have been preoccupied with questions about the origin and evolution of plastids and mitochondria, the DNA-containing organelles of eukaryotic cells. The first contribution to understanding the ancestry of photosynthetic eukaryotes was made by Schimper in 1883, who, based on his microscopic observation of different protists, postulated a prokaryotic origin of plastids (Schimper, 1883). In 1910, Mereschkowsky introduced the theory of symbiogenesis, according to which chloroplasts were once symbiotic cyanobacteria (Mereschkowsky, 1910). A few years later, Wallin proposed that mitochondria were also descendants of symbiotic bacteria (Wallin, 1927). Initially ignored by the science community, these theories were revived in the late 1950s and early 1960s as a consequence of the discovery that organelles carry their own genetic information (Nass and Nass, 1963; Stocking and Gifford Jr, 1959). In addition, electron microscopic analysis revealed that the chloroplast DNA fibrils resemble the naked DNA found in prokaryotes (Ris and Plaut, 1962). In 1967, Lynn Sagan-Margulis summarized all evidence in support of the endosymbiotic theory as the basis for the origin of eukaryotic cells (Sagan, 1967).

1.2.2 From an endosymbiont to an organelle

Nowadays, when the endosymbiotic theory is well accepted as the “truth” about the origin of plastids (and mitochondria, which will not be further discussed here), scientists’ efforts are directed towards unraveling when and how the evolution from an endosymbiont to an organelle occurred. Oxygenic photosynthesis is believed to have originated approximately 3.5 billion years ago with the appearance of cyanobacteria, the progenitors of plastids (Nisbet and Sleep, 2001). Structural and molecular phylogenitic analyses date plastid origin between 1.5 and 1.2 billion years ago. Initially, a eukaryote already carrying mitochondria, took up a free living cyanobacterium and became autotrophic (Dyall, et al., 2004). This event, referred to as primary endosymbiosis, led to the formation of three autotrophic lineages - chlorophytes/green algae

(giving rise to land plants), rhodophytes/red algae and glaucophytes. Later on, secondary endosymbiosis occurred, during which an autotrophic eukaryote became an endosymbiont of a heterotrophic eukaryotic cell, giving rise to euglenophytes, dinoflagelates, chryptogytes, etc (Gould, et al., 2008).

Since land plants originate from primary endosymbiosis, we will here mainly focus on the evolution of a free-living cyanobacterium to a cell organelle. Primary plastids have two envelope membranes, both derived from the Gram-negative cyanobacteria and no traces of a phagocytotic host membrane surrounding the plastids are detectable (Cavalier-Smith, 2000). Galactolipids and β-barrel proteins are characteristic features of both the outer envelope of plastids and the outer membrane of cyanobacteria (Gould, et al., 2008). The genome size reduction of the endosymbiont is the key event that drove the conversion of this free living species into a non-autonomous organelle. The vast majority of the cyanobacterial genes were lost or transferred to the nucleus of the host cell, while several genes mainly required for photosynthesis and gene expression were retained in the organellar genome (Kleine, et al., 2009). Nuclear copies of organelle genes acquired eukaryotic promoters and regulatory sequences in order to allow for nuclear expression, as well as targeting sequences needed for the shipping of their proteins to the appropriate organelle compartment (Martin and Herrmann, 1998). The DNA transfer from the organelle to the nucleus would thus not be possible without the presence of protein-import machineries that are localized in the two membranes of primary plastids. These transporters allow the plastid to take up nuclear encoded plastid protein precursors, cleave the transit peptide and release the mature form in the stroma (Strittmatter, et al., 2010). Moreover, the cyanobacterial endosymbiont not only contributed new biochemical features like photosynthesis and synthesis of starch to the host cell but also took over some of the preexisting host metabolic activities (Ball, et al., 2010; Gould, et al., 2008). Thus, it evolved into an organelle firmly integrated in the metabolism of the eukaryotic cell.

1.3 The plastid genome and it is organization

Plastids carry their own genetic information - a core set of genes retained from their cyanobacterial ancestor. The plastid genome (plastome) of land plants consists of 100-120 genes encoded in a 120-160 kilobase (kb) long circular double-stranded DNA molecule. (Figure 1;Green, 2011). In addition to monomeric circles, the plastome is found as circular multimers or

in various linear conformations *in vivo* (Lilly, et al., 2001; Oldenburg and Bendich, 2004, Oldenburg and Bendich, 2004). There are multiple plastome copies per plastid, with the number varying among species and tissues and during plastid differentiation (Baumgartner, et al., 1989; Isono, et al., 1997; Zoschke, et al., 2007). Several plastid genomes are packed together in nucleoprotein complexes (plastid nucleoids), which size, shape and numbers are also species dependent (Kuroiwa, 1991). The nucleoid is found anchored to the inner envelope or thylakoid membrane (Liu and Rose, 1992; Sato, et al., 1993). The transcription apparatus is believed to be tightly associated with the plastid genome, since isolated nucleoids retain transcriptional activity in vitro (Sakai, et al., 1991).

The complete genome sequences were first reported for tobacco and liverwort chloroplasts (Shinozaki, et al., 1986, Ohyama, et al., 1986). Now, 25 years later, more than 240 plastid genome sequences are available at the NCBI genome database (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid). This sequence information brought us better understanding of plastid genetics and genome evolution. Most plastid genomes have a quadripartite structure consisting of two inverted repeats (IRa and IRb) dividing the circle into a large (LSC) and small single copy (SSC) region (Figure 1). The two IRs are identical in sequence and thus the genes encoded in this regions, the ribosomal RNA genes and some additional ones, are present in two copies per genome. The exact reason for this duplication is still a matter of debate (Bock, 2007). The plastome has a low GC content (around 30-40 %) that is more pronounced in intergenic regions (Ohyama, et al., 1988). With more than 100 genes encoded only in 120-160 kb, the plastome is more densely packed than the nuclear and mitochondria genomes of the plant cell (Sugiura, 1992).

Plastid-encoded genes can be grouped in two main groups – photosynthesis genes and genetic system (housekeeping) genes (Figure 1), with the former one showing stronger conservation and higher GC content (Shimada and Sugiura, 1991). There are 47 photosynthesis-related genes in the plastome of flowering land plants (angiosperms), coding for subunits of the photosynthetic apparatus. Nevertheless, all the protein complexes involved in photosynthesis require additional nuclear-encoded components of cyanobacterial origin. The group of genetic system genes consists of 62 members involved in plastid gene expression (Bock, 2007). For example, functional *rpo* genes coding for homologues of the cyanobacterial RNA polymerase subunits α, β, β’ and β’’, which form the core of the plastid-encoded plastid RNA polymerase (Ohyama, et

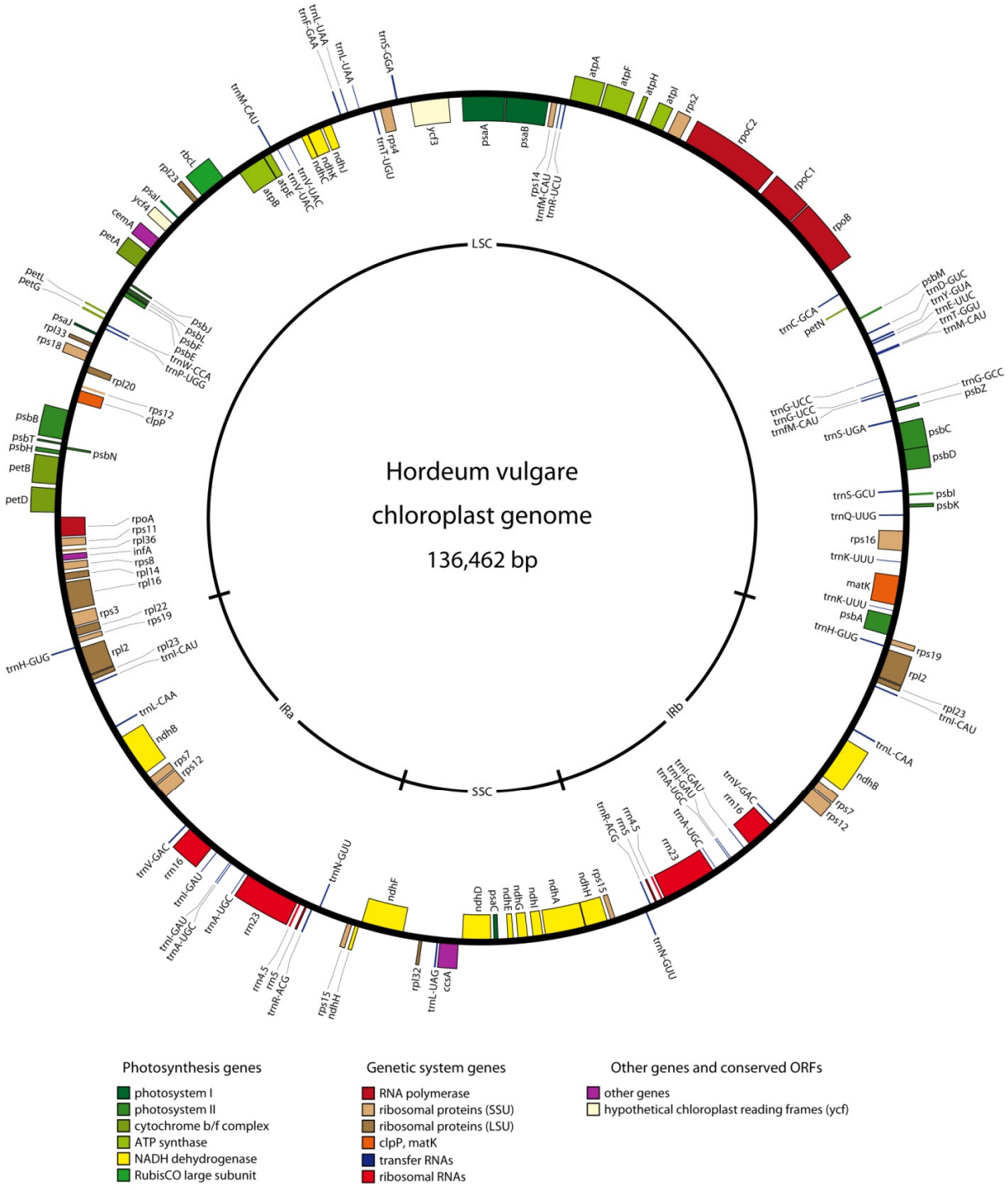


Figure 1: Physical map of barley (*Hordeum vulgare*) chloroplast genome. The inner circle depicts the quadripartite structure of the plastome. The outer circle shows the annotated genes. Genes at the inside and outside of this circle are transcribed clockwise and counterclockwise, respectively. The image was drawn using OrganellaGenomeDRAW (Lohse, et al., 2007) and further modified.

al., 1986; Shinozaki, et al., 1986). Additionally, tRNA and rRNA genes, as well as genes coding

for ribosomal proteins, an RNA maturase, a protease and the translation initiation factor IF-1 are encoded on the higher plant plastome (Hess, et al., 1994; Liere and Link, 1995; Sijben-Muller, et al., 1986; Shanklin, et al., 1995; Sijben-Muller, et al., 1986). While the plastome carries all the required tRNAs and rRNAs (Lung, et al., 2006), many additional proteins functioning in translation (e.g. two thirds of the ribosomal proteins) and transcription (e.g. sigma subunits of PEP and an additional nuclear-encoded plastid RNA polymerase) are encoded by nuclear genes and postranslationally imported in plastids (Peled-Zehavi and Danon, 2007; Liere and Börner, 2007). There are only a few conserved open reading frames (ORFs) with not yet known function or plastid-encoded genes which are not directly related to photosynthesis or gene expression (Bock, 2007).

1.4 Plastid transcription

Plastid biogenesis and differentiation are driven by the coordinated expression of nuclear and plastid genes. Regulation of gene expression in plastids can occur at the DNA level by adjusting the plastome copy numbers per organelle or through various transcriptional, post-transcriptional, translational and posttranslational control mechanism (Bendich, 1987, Bollenbach, et al., 2007; Liere and Börner, 2007, Schmitz-Linneweber and Barkan, 2007, Kanervo, et al., 2007; Marín-Navarro, et al., 2007). Plastid transcription regulation, even though initially regarded as having a minor effect on gene expression (Gruissem, et al., 1988), has been extensively studied the past years and now recognized as a molecular process with high degree of complexity and multiple levels of regulation.

1.4.1 Plastid RNA Polymerases

1.4.1.1 Plastid-encoded plastid RNA polymerase (PEP)

The plastid genome contains functional *rpo* genes which code for homologues of the cyanobacterial RNA polymerase subunits α, β, β' and β'', forming the core of the plastid-encoded plastid RNA polymerase (PEP; Figure 2; Ohyama, et al., 1986; Shinozaki, et al., 1986; Sijben-Muller, et al., 1986). Similar to the gene organization in bacteria, *rpoA*, coding for the α subunit of PEP, is found in a gene cluster together with several genes coding for ribosomal proteins (Purton and Gray, 1989), while *rpoB*, *rpoC* and *rpoC1,* encoding the β, β' and

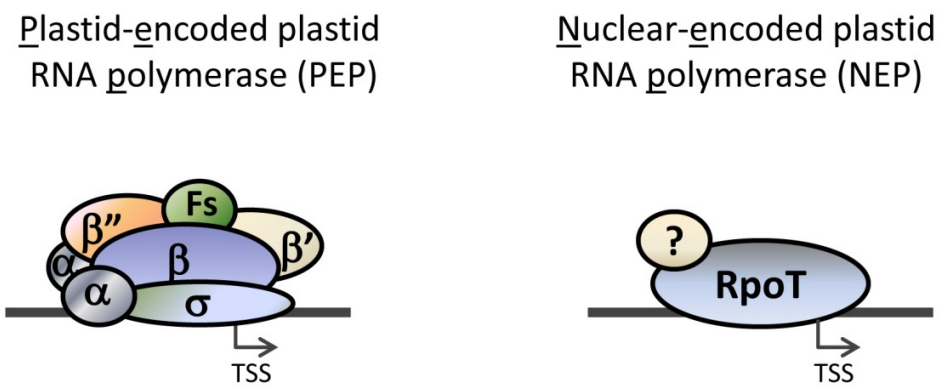


Figure 2: RNA polymerases in plastids. Transcription in plastid is dependent on two RNA polymerases: the plastid-encoded RNA polymerase (PEP) and the nuclear-encoded RNA polymerase (NEP). PEP is a multisubunit enzyme that is homologous to bacterial RNA polymerases. It consists of the plastid-encoded 2xα, β, β', and β" core-subunits and nuclear-encoded σ factors which confer promoter recognition. PEP can be found associates with additional nuclear-encoded factors (Fs). NEP is a single subunit phage-type RNA polymerase which may require additional, yet unknown transcription factors (?). The transcription start sites (TSSs) are indicated by arrows. Modified after Liere and Börner, 2007.

β''subunits, respectively, together form an operon (Shinozaki, et al., 1986, Kaneko, et al., 1996). PEP β and β' subunits can serve as functional substitutes of the homologous subunits of the *E.coli* RNA polymerase (Severinov, et al., 1996). Moreover, PEP is sensitive to tagetitoxin, an inhibitor of bacterial transcription (Mathews and Durbin, 1990), further demonstrating the high degree of conservation between the plastid encoded and eubacterial RNA polymerase. However, this evolutionary conservation did not allow for the substitution of the PEP α subunit with the *E.coli* homologue in transplastomic tobacco plants (Suzuki and Maliga, 2000).

PEP can be isolated from plastids as a soluble protein or an insoluble, DNA and other protein associated form, also known as "transcriptionally active chromosome" (TAC; Briat, et al., 1979; Krause and Krupinska, 2000). The soluble PEP fraction, isolated from etioplasts (photosynthetically inactive plastids) consists mainly of the core subunits (Pfannschmidt and Link, 1997). PEP preparations from photosynthetically active plastids are more complex, with the PEP holoenzyme found associated with various nuclear-encoded proteins required for transcription specificity and regulation under light conditions (Ogrzewalla, et al., 2002; Pfalz, et al., 2006; Pfannschmidt and Link, 1997; Pfannschmidt, et al., 2000). For example, among them are the sigma factors, which confer promoter recognition (Lerbs-Mache, 2011; Liu and Troxler, 1996; Schweer, et al., 2010; Tanaka, et al., 1996; Tanaka, et al., 1997). This functional

dependence of PEP on nuclear-encoded accessory proteins is a way to establish a nuclear control over plastid transcription.

1.4.1.2 Nuclear-encoded plastid RNA polymerase (NEP)

In stark contrast to the bacterial RNA polymerase, PEP is not sufficient to transcribe all plastid genes in higher plants. A second polymerase, denoted nuclear-encoded plastid RNA polymerase (NEP), was found to participate in transcription and be essential for plastid transcription (Figure 2; Allison, et al., 1996; Hess, et al., 1993; Siemenroth, et al., 1981). The first evidence for the existence of one or more NEP enzymes came from studies following the effect of inhibitors of translation on cytoplasmic and plastid ribosomes (Ellis and Hartley, 1971). Active RNA synthesis was detected in ribosome-deficient plastids, which implied a nuclear location of the gene(s) responsible for this activity (Bünger and Feierabend, 1980, Hess, et al., 1993; Siemenroth, et al., 1981; Han, et al., 1993). Moreover, transcription was detected in plastids of the parasitic plant *Epifagus virginiana*, which has a plastome that lacks genes for the core subunits of PEP (Ems, et al., 1995; Morden, et al., 1992). Similarly, plastid genes were found to be transcribed in transplastomic tobacco plants with knocked out PEP activity. However, these tobacco mutants had an albino phenotype, implying that NEP alone cannot provide for photosynthetically active chloroplasts (Allison, et al., 1996; Hajdukiewicz, et al., 1997; Legen, et al., 2002).

NEP is represented by one or more phage-type RNA polymerases in higher plants. *RpoT* (RNA polymerase of the phage T3/T7 type) genes coding for single-subunit RNAPs were discovered in many plant species (Liere and Börner, 2007). In dicots with diploid genomes, e.g. *Arabidopsis thaliana* and *Nicotiana sylvestris*, there are three *RpoT* genes – *RpoTm, RpoTp and RpoTmp*; their products are targeted to the mitochondria, plastids and both organelles, respectively (Figure 3; Hedtke, et al., 1997; Hedtke, et al., 2000; Hedtke, et al., 1999; Kobayashi, et al., 2001). Therefore, NEP activity in dicots can be carried out by two polymerases, RpoTp and RpoTmp. In monocots, with cereals being the only investigated family until now, NEP is represented by a single RpoTp polymerase (Chang, et al., 1999; Emanuel, et al., 2004). *Chlamydomonas*, a genus of green algae, possesses a single *RpoT* gene that is most likely coding for a mitochondrial RpoT and not a plastid-targeted form, since inhibition of the plastid-encoded RNA polymerase led to complete loss of plastid gene expression (Guertin and Bellemare, 1979;

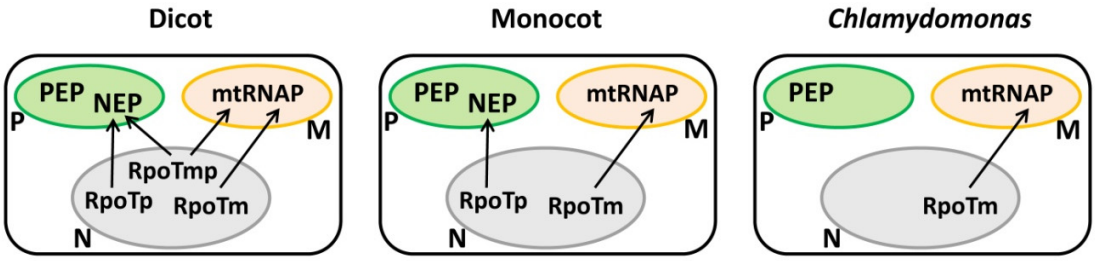


Figure 3: RpoT polymerases in organelles of different organisms. Genes in the nucleus (N) encode RpoT polymerases which are targeted to plastids (P) and/or mitochondria (M). In dicots with diploid genomes there are three *RpoT* genes – *RpoTm, RpoTp and RpoTmp*; their products are targeted to the mitochondria, plastids and both organelles, respectively. Therefore, NEP activity in dicots can be carried out by two polymerases, RpoTp and RpoTmp. In monocots. NEP is represented by a single RpoTp polymerase. *Chlamydomonas* possesses only one *RpoT* gene that is most likely coding for a mitochondria RNA polymerase (mtRNAP). Modified after Liere, et al., 2011.

Surzycki, 1969). There are multiple lines of evidence demonstrating that both RpoTp and RpoTmp are indeed responsible for the NEP activity in plastids. Both RpoTp and RpoTmp were detected in plastids using specific antibodies (Chang, et al., 1999, Azevedo, et al., 2006). Moreover, transgenic tobacco and *Arabidopsis* plants, overexpressing RpoTp, were characterized by an increased transcription from a set of NEP promoters (Liere, et al., 2004). Additionally, RpoTp was shown to recognize distinct NEP promoters *in vitro* (Kühn, et al., 2007). Even though RpoTmp was not able to bind NEP promoters *in vitro* (Kühn, et al., 2007), it was demonstrated that the enzyme plays a distinct role in plastid transcription at early developmental stages in *Arabidopsis* (Courtois, et al., 2007). Furthermore, knocking out *RpoTp* or *RpoTmp* genes in *Arabidopsis* resulted in plants with impaired chloroplast biogenesis and leaf morphogenesis, while *RpoTp/RpoTmp* double mutants exhibited an even more severe phenotype characterized by extreme growth retardation (Hricova, et al., 2006).

1.4.2 Architecture of plastid promoters

1.4.2.1 PEP promoters

So far, only a few plastid promoters have been mapped and analyzed in detail. However, it has become clear that PEP and NEP recognize distinct types of promoter sequences (Liere and Börner, 2007; Liere, et al., 2011). Due to the eubacterial origin of plastids, it is not surprising that the majority of promoters utilized by PEP are similar to *E. coli* σ70 promoters (Gatenby et

al., 1981; Gruissem and Zurawski, 1985; Strittmatter et al., 1985). Moreover, *E.coli* RNA polymerase is able to accurately transcribe from these PEP promoters (Boyer and Mullet, 1986; Bradley and Gatenby, 1985). PEP promoters consist of -35 (TTGaca) and -10 (TAtaaT) consensus elements (Gatenby, et al., 1981; Gruissem and Zurawski, 1985; Liere and Börner, 2007; Strittmatter, et al., 1985). Some PEP promoters are characterized by additional *cis*-regulatory sequences. For example, the mustard *psbA* promoter was shown to have an additional regulatory element (TATA-box) between the -10 and -35 promoter element, which was able *in vitro* to promote a basal level of transcription without the presence of the -35 region in plastid extracts from dark and light grown plants. Nevertheless, the -35 element was essential for reaching the full promoter activity required during active photosynthesis (Eisermann, et al., 1990; Link, 1984). In the case of the wheat *psbA* promoter, even though present, the TATA-box seems not to have an important function. Instead, an extended -10 sequence (TGnTATAAT) is utilized as the sole *psbA* promoter element by PEP isolated from the leaf tip which contains mature chloroplasts. PEP obtained from young plastids in the leaf base, however, still needed both the -10 and -35 boxes (Satoh, et al., 1999). Several *cis*-elements required for the binding of regulatory proteins have been also described. A twenty-two bp sequence, known as the AAG box, was found to play an important role in the regulation of the blue light-responsive promoter of *psbD* (coding for the photosystem II reaction center chlorophyll protein D2) by providing the binding site for the AGF (AAG-binding factor) protein complex, which acts as a positive regulator (Kim, et al., 1999). Similarly, the RLPB (*rbcL* promoter binding) factor was found to enhance transcription upon binding to the sequence -3 to -32 nt upstream of the *rbcL* transcription start site (Kim, et al., 2002).

1.4.2.2 NEP promoters

In green chloroplasts, PEP transcripts are overrepresented, while most of the transcripts generated by NEP are of low abundance, and are thus rarely detectable (Hess and Börner, 1999; Liere and Maliga, 2001). Therefore, identification of NEP transcription start sites has been only feasible in plants with knocked out/down PEP activity. Examples of such experimental systems are the *albostrians* barley and *iojap* maize mutants which carry ribosome deficient plastids, heterotrophically cultures BY2 cell of tobacco, transplastomic tobacco plants with deleted *rpo* genes, and *Arabidopsis* plants grown on spectinomycin which inhibits plastid translation

(Allison, et al., 1996; Hübschmann and Börner, 1998; Serino and Maliga, 1998; Silhavy and Maliga, 1998; Swiatecka-Hagenbruch, et al., 2007; Vera, et al., 1996).

The NEP promoters, analyzed so far, share high sequence similarity with phage and plant mitochondrial promoters (Allison, et al., 1996; Liere and Börner, 2007; Vera, et al., 1996). This is in agreement with the fact that RpoTm alone or together with RpoTmp transcribe the mitochondrial genome of plants and algae (Liere and Börner, 2007). Based on their architecture, NEP promoters can be grouped into three types (Weihe and Börner, 1999; Liere and Maliga, 2001). The majority of analyzed NEP promoters belong to Type-I NEP promoters, which are further divided into two subclasses. Type-Ia promoters consist of a conserved YRTa core motif located shortly upstream of the transcription start site (Liere and Börner, 2007; Liere, et al., 2011). A classical example of a Type-Ia NEP promoter is P*rpoB*-345 (transcription from this promoter initiates 345 nt upstream of the *rpoB* ORF in tobacco; Serino and Maliga, 1998). Deletion analysis of the 5'-flanking region of the *Arabidopsis rpoB* fused to GUS and transiently expressed in cultured tobacco cells suggested the existence of upstream regulatory elements in addition to the YRTa core (Inada, et al., 1997). On the contrary, no sequence elements outside of the core were found to have a significant influence on the *in vitro* transcription from the tobacco *rpoB* promoter (Liere and Maliga, 1999). Type-Ib NEP promoters are characterized by carrying an additional conserved sequence motif ($ATAN_{0-1}GAA$), referred to as the GAA-box, that is located approximately 18 to 20 bp upstream of the YRTA motif (Weihe and Börner, 1999; Liere and Börner, 2007). Deletion analysis of the tobacco P*atpB*-289 promoter revealed a functional role of this element in promoter recognition both *in vivo* and *in vitro* (Kapoor and Sugiura, 1999; Xie and Allison, 2002).

Transcription from Type-II NEP promoters is YRTa independent and is instead controlled by "non-consensus" promoter elements (Liere, et al., 2011). The most closely investigated example is the tobacco P*clpP*-53, with a regulatory core sequence found to comprise the region -5 to +25 with respect to the transcription initiation site (Sriraman, et al., 1998). Interestingly, the *clpP*-53 promoter sequence is conserved among monocots, dicots and *C. reinhardtii* but it does not drive transcription in rice and *Chlamydomonas*. However, when introduced into tobacco, the rice P*clpP*-53 sequence is efficiently utilized, which suggests that this promoter sequence might be recognized by a distinct transcription factor or a NEP enzyme that is present in dicots but not monocots, such as RpoTmp (Liere, et al., 2004; Sriraman, et al., 1998).

The Pc promoter of the *rrn* operon described in spinach, *Arabidopsis* and mustard represents another non-YRTa type of NEP promoters (Liere and Börner, 2007). The promoter region of the *rrn* operon is highly conserved in plants and contains both -10 and -35 PEP promoter elements, which are driving the transcription of the operon in barley, tobacco, maize and in later developmental stages of *Arabidopsis* (Allison, et al., 1996; Hübschmann and Börner, 1998; Strittmatter, et al., 1985; Vera and Sugiura, 1995; Courtois, et al., 2007). However, in spinach, as well as during the early development in *Arabidopsis*, NEP initiates at a site between the conserved PEP elements (Baeza, et al., 1991; Iratni, et al., 1994; Iratni, et al., 1997; Swiatecka-Hagenbruch, et al., 2007; Courtois, et al., 2007).

1.4.2.3 Internal promoters of tRNA genes

The majority of tRNA genes are transcribed by PEP from typical σ^{70}-like promoters upstream the transcription start site (Liere and Börner, 2007). However, there are reports suggesting transcription from internal promoters for several tRNA genes, *i.e*. the spinach *trnS*, *trnR* and *trnT* (Cheng, et al., 1997; Gruissem, et al., 1986), the mustard *trnS*, *trnH* and *trnR* (Liere and Link, 1994; Neuhaus and Link, 1990; Nickelsen and Link, 1990) and the *trnE* of *Chlamydomonas* (Jahn, 1992). Furthermore, the coding region of the *trnS* from spinach alone was shown to be sufficient to promote basal levels (8%) of transcription in *in vitro* assays (Wu, et al., 1997). However, the exact features of tRNA-internal promoter elements and the polymerase(s) recognizing them remain to be further elucidated.

1.4.3 General features of chloroplast transcription

Most chloroplasts genes are organized in complex operons, and thus are transcribed from upstream promoters into large polycistronic transcripts. Plastid operons are found conserved among plant species (Sugiura, 1992). Genes coding for subunits of a single complex or for proteins with common functions are often grouped in operons, which could facilitate their coordinated expression and stoichiometric accumulation. Moreover, this organization allows for the expression of housekeeping genes independently of photosynthesis genes (Mullet, 1993). However, there are still operons, e.g. the *psbB* and *ndhD* operons, carrying genes with unrelated functions. The differential expression of several genes within operons is ensured by multiple posttranscriptional regulatory steps (see Chapter 1.5.; Barkan, et al., 1994; Del Campo, et al.,

2002; Felder, et al., 2001; Hirose and Sugiura, 1997; Westhoff, 1985). Another mechanism that allows for the differential expression within operons is the usage of promoters internal to transcription units. For example, several internal promoters were detected within the *psbK* operon to permit the synthesis of smaller polycistronic units (Sexton et al., 1990a; Sexton, et al., 1990).

Another general feature of plastid transcription is the usage of multiple promoters, leading to a variety of transcripts generated per gene (Liere and Börner, 2007). However, the exact role of only a few of the multiple promoters has been investigated. The blue-light-responsive promoter (BLRP) upstream of *psbD* is among the best studied cases. BLRP is one of the three PEP promoters driving the transcription of *psbD* and *psbC*, coding for the reaction center protein D2 and the chlorophyll-binding antenna protein CP43 of photosystem II, respectively (Sexton, et al., 1990; Christopher, et al., 1992; Kim and Mullet, 1995). BLRP was found to maintain high transcription rates of *psbD* and *psbC* in mature chloroplasts, and thus allowing for the re-synthesis and replacement of D2 and CP43 (Sexton, et al., 1990), which are damaged or degraded in illuminated plants (Mattoo, et al., 1989).

1.4.4 Division of labor between PEP and NEP in plastid transcription

Over the past 20 years, many studies have been dedicated to unravel the function and interplay of PEP and NEP in plastid transcription. A complex picture has emerged, with PEP and NEP abundance, transcriptional activity, promoter usage, and transcript stability varying significantly among species, tissue and plastid types, and during plastid development (Liere and Börner, 2007; Liere, et al., 2011). Even though several models describing their distinct role in organellar transcription have been proposed, the exact division of labor between PEP and NEP remains poorly understood.

A clue to the function of PEP and NEP in plastid transcription could be provided by investigating the distribution of promoters recognized by these two enzymes. So far, only a few plastid promoters have been mapped and further analyzed. According to the current view, most genes coding for house-keeping proteins have both PEP and NEP promoters, photosynthesis genes have only PEP promoters and a few house-keeping genes are transcribed exclusively from NEP promoters (Hajdukiewicz, et al., 1997; Swiatecka-Hagenbruch, et al., 2007). Interestingly, the *rpoB* operon, coding for three of the four subunits forming the core of PEP, is solely

transcribed by NEP (Hübschmann and Börner, 1998; Silhavy and Maliga, 1998). In this way, the presence and abundance of PEP is dependent on NEP, and thus tightly controlled by the nucleus.

It was proposed that transcription of PEP and NEP, through recognition of distinct promoters, can serve as a general mechanism of group-specific gene regulation during chloroplast development (Hajdukiewicz et al., 1997). According to this model, in proplastids, NEP plays an essential role in chloroplast gene expression by transcribing housekeeping genes and the *rpoB* operon, while in chloroplasts, PEP starts transcribing photosynthesis genes and takes over the transcription of housekeeping genes. Indeed, NEP promoters were found to be more active in early leaf development, while transcription activity of PEP is reported to increase during chloroplast maturation (Baumgartner, et al., 1993; Courtois, et al., 2007; Demarsy, et al., 2011; Emanuel, et al., 2004; Kapoor, et al., 1997; Swiatecka-Hagenbruch, et al., 2008; Zoschke, et al., 2007). For example, the transcription of the *rrn* operon in *Arabidopsis* during seed germination and early plant development was shown to be carried out by NEP (RpoTmp), with PEP taking over this task in later developmental stages (Courtois, et al., 2007). However, transcripts generated by NEP were found to cover the entire plastome in leaves of transplastomic tobacco plants lacking PEP activity (Krause, et al., 2000, Legen, et al., 2002). Even though this observation was mainly attributed to spurious transcription initiation by NEP throughout the plastome, it could as well indicate a more general function of NEP in chloroplast transcription.

Cahoon et al. (2004) investigated plastid transcription in the leaf base (proplastids) and leaf tip (chloroplasts) and proposed a model for PEP-NEP dynamics in maize. On one hand, NEP was observed to become less abundant as chloroplasts mature and this was correlated with both an increase in NEP transcriptional activity and decrease in the stability of NEP transcripts. This would result in no or little accumulation of NEP-controlled transcripts during plastids development. On the other hand, transcription rates of PEP increased during chloroplasts development, with the RNA stability remaining constant or even increasing, and thus leading to an overall increase in the PEP-controlled transcript accumulation in mature chloroplasts. However, such a strong correlation between polymerase usage and polymerase-specific transcript accumulation was not observed in *Arabidopsis* (Zoschke, et al., 2007).

In *Arabidopsis*, the activation of a NEP promoter was described to compensate for abolished transcription from the *atpB* PEP promoter (Schweer, et al., 2006). Thus, in certain cases, NEP may function as an SOS-enzyme in plastid transcription. Moreover, it was proposed that by

acquiring a plastid localization signal and being able to recognize promoters with simple structure, the nuclear encoded RpoTs might have functioned in compensating for the degeneration of PEP promoters. Thus, the complexity of plastid transcription may have evolved to guarantee functional chloroplast gene expression (Maier, et al., 2008).

1.5 Plastid RNA processing

Similar to transcription, plastid RNA processing represents an intricate combination of prokaryotic and eukaryotic features of gene expression. Like bacteria, chloroplasts express the majority of their genes from operons, leading to the synthesis of long polycistronic RNAs. However, the bacterial concept of the operon, as a cluster of coregulated set of genes (Jacob and Monod, 1961), does not fully apply for plastids. Instead of being directly translated, numerous polycistronic transcripts are rather functioning as precursors, excessively cleaved into smaller polycistronic or monocistronic RNAs, with many of these species still requiring splicing and/or RNA editing (which will not be further discussed here) to become functional (Barkan, 2011; Stern, et al., 2010). Thus, in addition to transcription from multiple promoters, RNA maturation further increases the complexity of RNA populations arising from most genes. Major events in plastid RNA maturation, *e.g.* 5' and 3' processed end formation and intercistronic processing, involve the action of ribonucleases with low sequence specificity, with the extent of processing determined by barriers like RNA-binding proteins and secondary structures.

1.5.1 Pentatricopeptide repeat proteins

The machinery of organellar post-transcriptional processes consists of numerous nuclear encoded chloroplast RNA-binding proteins. The PPR (pentatricopeptide repeat) proteins comprise a superfamily of helical repeat proteins in higher plants (Schmitz-Linneweber and Small, 2008). They are defined by loosely conserved helical repeats comprising 35 amino acids and have been shown to function almost exclusively in organellar gene expression (Saha, et al., 2007; Small and Peeters, 2000). The current view is that most PPR proteins have sequence-specific single stranded RNA-binding activity. PPR proteins are involved in variety of posttranscriptional processes, e.g. editing, splicing, RNA stability control. For example, the accumulation of chloroplast RNAs with processed 5' and 3' ends mapping in certain intergenic regions was shown to be dependent on the PPR proteins CRP1, PPR10 and HCF152 (Barkan, et

al., 1994; Fisk, et al., 1999; Meierhoff, et al., 2003; Pfalz, et al., 2009). Recently, the underlying mechanism of PPR10 role in mRNA maturation, by serving as site specific barrier to nucleases, has been elucidated (Figure 4; Pfalz, et al., 2009; Prikryl, et al., 2011, Barkan, 2011). Moreover, other PPR and "PPR-like" (other RNA binding proteins with helical repeat architecture) were shown to facilitate the stabilization of several chloroplast 5' RNA termini (Johnson, et al., 2010; Loiselay, et al., 2008; Vaistij, et al., 2000; Yamazaki, et al., 2004; Sane, et al., 2005).

1.5.2 Plastid ribonucleases involved in 5' and 3' RNA maturation

Both endo- and exonuclease activities, mediated by nuclear-encoded ribonucleases (RNases), have been reported to participate in maturation of rRNAs and tRNAs, intercistronic mRNA processing, and RNA decay in plastids (Barkan, 2011; Bollenbach, et al., 2007; Stern, et al., 2010). It is currently believed that many plastid RNases are homologous to bacterial ribonucleases. However, in many cases, the enzymes and their precise function have not been elucidated (Stoppel and Meurer, 2011). Among the best characterized plastid ribonucleases are the RNases participating in 5' and 3' RNA maturation.

Processed 5' RNA ends have been hypothesized to emerge either *via* a 5'-3' exonuclease pathway or an endonucleolytic cleavage (Stern, et al., 2010). Homologues of *E.coli* RNase E and *B. subtilis* RNase J are suggested to act as major plastid endonucleases. *Arabidopsis* RNase E was shown to act in a similar manner as its *E.coli* counterpart: it prefers 5' monophosphorylated (processed) substrates; is inhibited by structured RNA; and preferentially cleaves AU-rich sequences (Mudd, et al., 2008; Schein, et al., 2008). Recently, *Arabidopsis* RNase J was also demonstrated to exhibit endonucleolitic activity but is insensitive to the number of phosphates at the 5' end. Moreover, similar to its *B. subtilis* homologue, plastid RNase J could also act as a 5' to 3' exonuclease with a preference to 5'-monophosphorylated RNAs (Sharwood, et al., 2011). RNase E and J endonucleolitic activity has been proposed to initiate intercistronic mRNA processing, and RNase J to further mediate the 5' to 3' trimming of RNAs, being blocked by RNA-binding protein barriers (see Chapter 1.5.3.; Barkan, 2011). Moreover, RNase J is suggested to act as surveillance enzyme that eliminates long asRNAs resulting from read-through transcription (Sharwood, et al., 2011).

In plastids, similar to bacteria, PNPase, a polynucleotide phosphorylase, participates in RNA processing, polyadenylation and degradation (Bollenbach, et al., 2007; Germain, et al., 2011;

Stern, et al., 2010). It was shown that PNPase catalyzes both processive 3' to 5' degradation and RNA polymerization (Yehudai-Resheff, et al., 2001). PNPase is suggested to act as a major 3' to 5' exonuclease generating processed mRNAs 3' termini (Walter, et al., 2002). The model of 3' mRNA maturation in plastids involves the presence of either a stable secondary structure, *e.g.* a stem-loop structure, or a protein bound to the RNA, which will serve as a barrier to the PNPase activity (Barkan, 2011; Stern, et al., 2010). Maturation of rRNAs is suggested to involve another 3' to 5' exonuclease called RNase R (Bollenbach, et al., 2005), while tRNA maturation is based on the endonucleases RNaseP and RNase Z, which act at the 5' and 3' end, respectively (Canino, et al., 2009; Schiffer, et al., 2002; Thomas, et al., 2000; Wang, et al., 1988).

1.5.3 Intercistronic mRNA processing

Plastid RNA metabolism is characterized by excessive intercistronic mRNA processing, *i.e.* processing of polycistronic transcripts between the coding regions. Initially, it was considered that intercistronic processing is mediated by site-specific endonucleases which generate processed 5' and 3' ends mapping to adjacent nucleotides (Bollenbach, et al., 2007). However, it was observed that the 5' processed end of *petD* and the 3' one of the upstream gene (*petB*) overlapped approximately 30 nt, and thus could not have been generated by a single cleavage event (Barkan, et al., 1994). A similar phenomenon was detected for other adjacent processed RNAs in maize. A detailed analysis of the processed termini mapping to the *atpI-atpH* and *psaJ-rpl33* intergenic regions led to the emergence of a model in which the maize PPR10, binds to these intergenic regions, and by serving as a barrier to 5' to 3' and 3' to 5' exonuclease activity defines the corresponding 5' and 3' processed plastid ends, respectively (Figure 4; Pfalz, et al., 2009). Indeed, recombinant PPR10 was found to be sufficient to block 5' to 3' and 3' to 5' exonuclease *in vitro*. Moreover, PPR10, supplemented with a generic 5' to 3' exonuclease, was shown to generate a 5' end that matches precisely the PPR10-dependent terminus generated *in vivo* (Prikryl, et al., 2011). Furthermore, three other PPR proteins, CRP1, CHF152 and PPR38, were shown to mediate the accumulation of RNAs with processed 5'/3' termini mapping to intergenic regions (Barkan, et al., 1994; Hattori and Sugita, 2009; Meierhoff, et al., 2003). Taken together, these observations hint for a contribution of PPRs (or other helical repeat proteins) in 5' and 3' processed end formation *via* binding to target RNA, and thus protecting adjacent regions by acting as a barrier to exonucleases. It was predicted that such an event should be accompanied

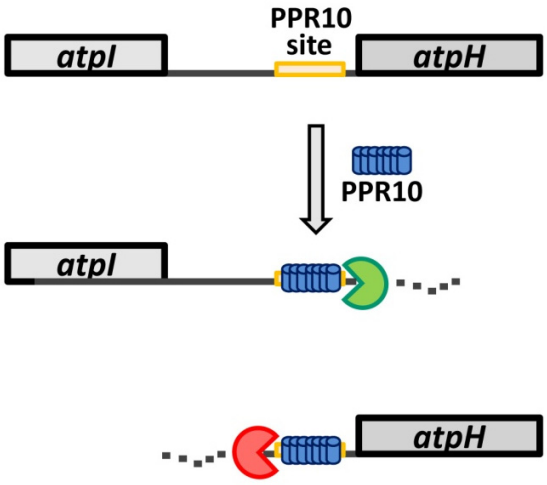


Figure 4: PPR10-dependent intercistronic processing. PPR10 binds specifically to *atpI-atpH* intergenic region (PPR10 binding site indicated) and by serving as a barrier to 5' to 3' (red packman) and 3' to 5' (green packman) exonuclease activity defines the corresponding 5' and 3' processed transcript ends, respectively.

by the presence of short RNA fragments representing PPR "footprints" *in vivo*– the minimal PPR binding sites, i.e. the regions protected from complete elimination by nucleases (Pfalz, et al., 2009). Indeed, small RNAs (sRNAs) corresponding to the PPR10's binding sites were detected in the sRNA transcriptome of several angiosperms (Johnson, et al., 2007; Morin, et al., 2008; Pfalz, et al., 2009; Ruwe and Schmitz-Linneweber, 2011; Zhelyazkova, et al., 2011).

1.6 Non-coding RNA in plastids

Non-coding RNAs (ncRNAs) play important roles in fine-tuning gene expression in both prokaryotes and eukaryotes (Lioliou, et al., 2010; Prasanth and Spector, 2007). In eukaryotes, posttranscriptional gene silencing by micro and small-interfering RNAs (18-25nt) functions as a major regulatory mechanism (Fire, 1999; He and Hannon, 2004). In bacteria, there are multiple examples of both *cis*- (e.g. riboswitches; Winkler and Breaker, 2005) and *trans*-acting regulatory RNAs (e.g. small RNAs; Papenfort and Vogel, 2009). The majority of bacterial *trans*-acting ncRNAs function as antisense RNAs (asRNA) by base paring to mRNAs and thus modifying their stability and/or translational efficiency. asRNAs can be *cis*- (bind and regulate the complementary sense RNA) or *trans*-encoded (regulate one or more loci *via* short regions of complementary), and are generated by transcription of free-standing genes rather than processing

(Lioliou, et al., 2010; Repoila and Darfeuille, 2009). Here, the terms *cis*/*trans*-encoded ncRNA and *cis*/*trans*-encoded asRNA will be used interchangeably.

So far, there are only a few reports describing the occurrence of non-coding RNAs in chloroplast. Hegeman et al. (2005) first hypothesized potential regulatory function of ncRNA in plastids, by describing the interference of an artificial antisense RNA with the editing of the corresponding sense RNA (Hegeman, et al., 2005). Later on, one study identified 12 ncRNAs in tobacco chloroplasts (Lung, et al., 2006). However, none of them was proven to originate from independent RNA genes, rather than being products of processing or degradation of read-through transcripts. Recently, a long asRNA to *ndhB* was proposed to be involved in stability control and RNA maturation of the complementary transcript (Georg, et al., 2010). Another asRNA was suggested to have an effect on the processing and accumulation of the 5S rRNA (Hotto, et al., 2010; Sharwood, et al., 2011). Furthermore, *cis*-encoded asRNAs were observed to form RNA-RNA hybrids with *psbT* mRNAs, and thus proposed to function in *psbT* translational inactivation by blocking the access to its ribosomal binding site (Zghidi-Abouzid, et al., 2011).

The above mentioned reports focus on single RNAs and do not point to non-coding RNA synthesis as a general feature of plastids. Cyanobacteria, the chloroplast progenitors, were shown to possess a plethora of ncRNAs for regulation of their gene expression (Georg and Hess, 2011). Therefore, it is highly possible that chloroplasts may contain additional hitherto not detected genes for potentially regulatory ncRNAs. Most recent studies indeed suggest that non-coding RNAs are common in the chloroplast transcriptome (Demarsy, et al., 2011; Hotto, et al., 2011; Mohorianu, et al., 2011; Wang, et al., 2011). Yet, these studies could not discriminate between ncRNAs generated via transcription rather than processing or degradation.

1.7 dRNA-seq - a powerful tool for mapping both primary and processed 5' ends

With the emergence of "next-generation" sequencing technologies, that parallelize the sequencing process, it became feasible and affordable to use DNA sequencing for various experimental applications, *e.g.* whole-genome sequencing, metagenomics, transcriptome sequencing, etc. RNA-seq, also referred to as massively parallel cDNA sequencing or whole transcriptome shotgun sequencing, has been revolutionizing global transcriptomic analysis by providing invaluable insights into the RNA populations and gene expression patterns of both prokaryotes and eukaryotes (Pinto, et al., 2011; Ozsolak and Milos, 2011).

Differential RNA-sequencing (dRNA-seq) is a recently established method, designed to selectively identify primary transcripts (Sharma, et al., 2010). It has already proven to be a powerful tool for mapping transcription start sites (TSSs) and to identify ncRNAs in several bacterial and archaeal species (Jäger, et al., 2009; Mitschke, et al., 2011; Sharma, et al., 2010). The method is based on the comparison of Terminator exonuclease (TEX) treated (TEX +) and non-treated (TEX -) RNA samples. TEX degrades RNAs with a 5' monophosphate (i.e. processed transcripts), but not with a 5' triphosphate or 5' CAP structure (i.e. primary, unprocessed transcripts). The comparison of cDNA libraries generated from TEX- and TEX+ samples can therefore be exploited to identify the protected primary transcripts and their TSSs. The phosphorylation status of processed 5' ends of plastid transcripts has not been directly investigated, but enzymes homologous to bacterial processing RNases are suggested to be involved in the processing of chloroplast transcripts (Stern, et al., 2010; Walter, et al., 2010). Moreover, the T4 ligase, which requires 5' monophosphates for its activity, ligates oligonucleotides to processed plastid transcripts (Swiatecka-Hagenbruch, et al., 2007). Hence, 5' ends of processed chloroplast transcripts most likely also have a monophosphate while the primary transcripts carry a triphosphate, as in bacteria. Therefore, it should be possible to distinguish and map both primary and processed RNA 5' termini in chloroplasts using TEX-based dRNA-seq.

1.8 The aim of this work

The current view on transcription regulation and RNA maturation in plastids is mainly based on the analysis of a few individual transcripts. Therefore, it hitherto remains rather unclear if the observations made are rules rather than exceptions. The present study aims to get a deeper insight into the plastid transcriptome. Here, a novel differential RNA-seq approach (dRNA-seq), designed to discriminate between prokaryotic primary and processed transcripts, will be used for the first to time to catalogue plastid RNA species and unravel the transcriptional organizations of genes in green and white plastids of the *Hordeum vulgare* (barley) line *albostrians*. Green *albostrians* plastids are phenotypically identical to wild-type chloroplasts, and thus their transcription relies on both PEP and NEP activity. In this study, they will be regarded to as an equivalent of mature barley chloroplasts. On the other hand, white *albostrians* plastids are

ribosome deficient and lack all plastid-encoded proteins including the core subunits of PEP. Therefore, transcription in these mutant plastids is carried out exclusively by NEP.

A key aim of this work will be to deliver a broader view of the division of labor between PEP and NEP. A clue to their function in plastid transcription could be provided by a global investigation of the distribution of promoters recognized by these two enzymes. Here, the transcription start sites in green (transcription by PEP and NEP) and white (transcription by NEP) *albostrians* plastids will be mapped and analyzed. General features of chloroplasts transcription, as well as polymerase specific gene expression in mature plastids will be investigated on a genome wide scale. Moreover, the regions upstream the transcription start sites will be analyzed for conserved promoter elements.

Furthermore, this study will address the question of the prevalence of ncRNAs encoded by free-standing genes in mature chloroplasts. Several studies have already suggested that ncRNAs are common in chloroplasts. However, they did not directly detect ncRNAs generated via transcription, the so far most abundant class of known regulatory ncRNAs in bacteria. Given the prokaryotic origin of plastids, it is worth screening the transcriptome of barley chloropalsts for the existence of ncRNAs generated by transcription initiation.

PPR and PPR-like proteins are involved in RNA processing and stabilization. Recently, PPR 10 was shown to participate in the processed termini formation of several plastid mRNAs by acting as a barrier to nucleases. However, the global impact of this mechanism on the chloroplast transcriptome has not yet been demonstrated. An assumption based on the current model for intercistronic mRNA processing is the presence of small RNAs representing *in vivo* “footprints” of bound PPR proteins. Last but not least, this work will investigate the occurrence of such small RNAs in the transcriptome of mature barley chloroplasts.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and Providers

Chemicals used in this study were purchased from Biozym, Merck, Carl Roth, Serva and Sigma-Aldrich unless specified otherwise. Ultrapure water was obtained using a USF Purelab Plus system. Sterilization of solutions and inactivation of genetically modified material was done for 20 min at 120 °C/ 55 kPa using Varioklav 75 S steam sterilizer (Thermo Scientific). The following providers were used:

Ambion, Applied Biosystems ,Invitrogen	Life Technologies GmbH, Darmstadt, Germany
Bio-Rad	Bio-Rad Laboratories GmbH, Munich, Germany
Biozym	Biozym Scientific GmbH, Hessisch Oldendorf, Germany
Calbiochem, Merck	Merck KGaA, Darmstadt, Germany
Epicentre	Epicentre Biotechnologies, Madison, WI, USA
Eurofins MWG Operon	Eurofins MWG GmbH, Ebersberg, Germany
Fermentas, Thermo Scientific	Thermo Fischer Scientific, Waltham, MA, USA
GE Healthcare	GE Healthcare Deutschland, Munich, Germany
Metabion	Metabion GmbH, Martinsried, Germany
QIAGEN	QIAGEN GmbH, Hilden, Germany
Promega	Promega GmbH, Mannheim, Germany
Carl Roth	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Serva	SERVA Electrophoresis GmbH, Heidelberg, Germany
Sigma-Aldrich; Sigma Life Science	Sigma-Aldrich Chemie GmbH, Steinheim, Germany

2.1.2 Oligonucleotides

DNA oligonucleotides were designed using Primer3 (http://frodo.wi.mit.edu/primer3/; Rozen and Skaletsky, 2000) and obtained from Sigma Life Science or Eurofins MWG Operon. The RNA linker used in 5'-RACE analysis was provided from Metabion. Sequences of the nucleotides used in in this work are provided in the respective Methods chapters.

2.1.3 Plant material

The barley mutant line *albostrians* (*Hordeum vulgare* L. cv. 'Haisa') was used as a model organism in this study. The progeny of homozygous *albostrians* plants consists of green, white and striped seedlings in a ratio of approximately 1:1:8 (Hess, et al., 1993). The first leaves from completely green and white *albostrians* seedlings were harvested and used for plastid isolation.

2.1.4 Bacterial strains

Plasmids containing 5'-RACE, 3'-RACE or cRT-PCR inserts were propagated in *E. coli* TOP10 cells (Invitrogen).

2.2 Methods

2.2.1 Plant growth

The barley mutant line *albostrians* was grown for 11 days in soil at 23 °C in a growth chamber with a photoperiod of 16h (light intensity: 150 μE s^{-1} m^{-2}).

2.2.2 Plastid isolation

The first leaves from completely green and white *albostrians* seedlings were harvested and used for plastid isolation as previously described (Zubo, et al., 2008). 10 μg of green and white leaf material were homogenized in 90 mL of homogenization buffer containing 0.33 M Sorbitol, 50 mM Tricine, pH 8.0, 2 mM EDTA, and 5 mM β-mercaptoethanol. The homogenate was filtered through two layers of Miracloth (Calbiochem-Behring) and centrifuged at 4,000 rpm (green) or 10,000 rpm (white) for 10 min. The pellet was resuspended in 1.5 mL of homogenization buffer and fractionated in a 30% / 70% (green) or 10% / 20% / 70% (white) discontinuous Percoll (GE Healthcare) gradient by centrifugation at 6,000 rpm (green) or 12,000 rpm (white) for 30 min. Intact chloroplasts from green leaves were collected at the interphase between 30% and 70% of Percoll. Plastids from white leaves were collected at the interphase between 20% and 70% (intact plastids) and 10% and 20% (semi-intact plastids) and a mixture of both intact and semi-intact plastids was used in the subsequent steps. All procedures were performed at 4 °C. Isolated plastids were washed two times with homogenization buffer, pelleted and stored at -80 °C until further used for RNA extraction.

2.2.3 Isolation of ribonucleic acids

2.2.3.1 Isolation of total RNA from green and white albostrians plastids

Total RNA was extracted from isolated green and white plastids using TRIzol (Invitrogen) following the manufacturer's protocol.

2.2.3.2 Determination of RNA concentration

The DNA and RNA concentration was determined by optical density with a spectrophotometer at 260 nm (Nanodrop; Thermo Scientific) and the integrity of rRNA bands was additionally verified by electrophoresis on 1% denaturating agarose gels containing 1.7 M formaldehyde. The RNA was stored at -80 °C.

2.2.4 Gel electrophoresis of nucleic acids

Agarose gel electrophoresis of DNA and RNA were performed as previously described (Sambrook and Russell, 2001).

2.2.5 Polymerase chain reaction (PCR)

PCR reactions were set up using *Taq* DNA Polymerase (QIAGEN) following the manifacturer's protocol. Nested PCRs used in 5'-RACE, 3'-RACE and cRT-PCR analysis were carried out at 56 °C and 58 °C annealing temperature in the first and second PCR, respectively. PCR products were analyzed by agarose gel electrophoresis.

2.2.6 RNA-seq

2.2.6.1 Depletion of processed transcripts, cDNA library preparation and 454 sequencing

Depletion of processed RNAs was performed as previously described (Sharma, et al., 2010). In brief, total RNA from green and white plastids was first cleaned from genomic DNA contamination by gDNA Wipeout buffer (QIAGEN). For depletion of processed transcripts, 7 µg of RNA from each sample was treated with Terminator^{TM} 5'-phosphate-dependent exonuclease (TEX; Epicentre) or in buffer alone for 60 min at 30 °C. 1 unit TEX was used per 1 µg total chloroplast RNA. Following organic extraction (25:24:1 v/v phenol/chloroform/ isoamyalcohol), RNA was recovered by overnight precipitation with 2.5 volumes of ethanol/0.1M sodium acetate

(pH 6.5). RNA was further treated with 1 unit tobacco acid pyrophosphatase (TAP; Epicentre) for 1 hour at 37 °C to generate 5'-mono-phosphates for linker ligation, and again purified by organic extraction and precipitation as described above. cDNA library preparation and 454 pyrosequencing were performed as previously described (Berezikov, et al., 2006) but omitting size fractionation. Briefly, equal amounts of +/- TEX treated RNA from green and white plastids was poly A-tailed using poly(A) polymerase followed by RNA linker ligation to the 5' phosphate of the RNA. First strand cDNA synthesis was performed using oligo(dT) primer and M-MLV-RNase H- reverse transcriptase. cDNA was PCR-amplified using a high fidelity RNA polymerase and linker-specific primers. Four cDNA libraries were generated: G+ (total RNA from green plastids treated with TEX); G- (total RNA from green plastids not treated with TEX); W+ (total RNA from white plastids treated with TEX) and W- (total RNA from white plastids not treated with TEX). Each library had a specific 4-mer barcode sequence attached to the 5' end of the cDNAs during the PCR amplification step. Sequencing was performed on Roche 454 FLX machines at the MPI for Molecular Genetics (Berlin, Germany).

2.2.6.2 Read mapping

For mapping of the cDNAs to the barley chloroplast genome, 5'-linker and polyA-tail clipped reads of at least 18 nt were aligned to the available sequence in NCBI (NC_008590) using WU Blast 2.0 (http://blast.wustl.edu/) with the following parameters: -B=1 -V=1 -m=1 -n=-3 -Q=3 -R=3 -gspmax=1 -hspmax=1 -mformat=2 -e=0.0001.

2.2.6.3 Data visualization

For each library, graphs representing the number of mapped reads per nucleotide were calculated and visualized using the Integrated Genome Browser (IGB) version 6.1 software from Affymetrix (http://genoviz.sourceforge.net/) as previously described (Sittka, et al., 2008). The graphs were normalized to the total number of mapped reads in each library and the y-axis indicates per mill mapped reads at a given position.

2.2.7 Annotation of TSSs and PSs based on dRNA-seq

A 5’ end was considered a TSS when it was both represented by at least two cDNA reads starting at the same genomic position (nucleotide) and found enriched after the TEX treatment. A 5’ end represented by one cDNA was considered only as a gTSS when it extended into an annotated gene and a consensus promoter was predicted upstream. A 5’ end represented by less cDNAs in (+) than (-) green libraries (see Appendix A-1, Column: *Not enriched*) is still accepted as a TSS if: (i) it agreed with already published data; or (ii) was verified as a TSS by 5’-RACE (Appendix B); or (iii) it was enriched after TEX treatment in white libraries. The 5’ ends of *trn* transcripts mapped to +1 relative to gene start were considered as TSSs only if they were found at least to be twofold enriched in (+) vs. (-) libraries. Stepwise accumulation of cDNAs enriched in (+) libraries with similar 3’ ends mapped shortly downstream of annotated ORFs were regarded as signals corresponding to 3’ UTRs. Several enriched 5’ ends mapped within five consecutive nucleotides were considered a single TSS and denoted by the genomic position of the most abundant of the 5’ ends in (+) libraries. The less abundant 5’ ends are listed in the Comments section of Appendix A. gTSSs with cDNA reads that did not reach into the corresponding downstream gene are referred to as “disconnected”.

All 5’ ends (except mature *trn* and *rrn* ends) represented by at least 5 cDNAs in (–) libraries and found significantly underrepresented after the TEX treatment were listed as processing sites (PSs). PSs represented by cDNAs that do not reach the downstream located gene, are referred to as “disconnected” (Appendix G). PSs mapped within 10 consecutive nucleotide were considered a single 5’ end and denoted by the genomic position of the most abundant of the 5’ ends in the (-) library. The less abundant 5’ ends are listed in the Comments section.

2.2.8 Promoter analysis

The -1 to -25, -26 to -50, -1 to -10 and -50 to +25 nt regions of 176 and 244 TSS mapped by dRNA-seq (Appendix A) in green and white plastids, respectively, were screened for common motifs with length of 3 to 9 nucleotides using MEME (Bailey, et al., 2009) version 4.7.0 with default cut-off values. MEME outputs were corrected for the expected distances of the detected motifs in respect to +1 (Liere and Börner, 2007). The occurrence of detected MEME motifs was further examined using MAST (part of the MEME suite). A Pearson's Chi-squared test with Yates' continuity correction of MAST scores was used to determine if the motifs were found

significantly overrepresented in the analyzed sequences of green or white plastids (p-value < 0.05 = significant). The -10 PEP motif (Figure 10A) was detected using MAST in the -1 to -25 region of 71 of the 176 promoter sequences of green TSS and in 51 of the 244 promoter sequences of the white TSS, which resulted in a p-value = 2.434e-05. For the PEP -35 consensus region found in the -26 to -50 nt of the green plastid TSS (Figure 10A) 22 MAST hits were found in green and 12 in white sequences leading to a p-value of 0.008552. The motif within -1 to -25 nt region of white TSS (Figure 10C, right) was found using MAST 38 times in white pre-TSS sequences and 12 times in pre-TSS sequences of green plastids (p-value = 0.009847). The -26 to -50 nt region of white TSS appeared to be highly variable and no motif could be identified in this region. The motif found within the first 10 nt upstream of the TSS in white plastids (Figure 10C, left) was found by MAST in 75 of the white and 22 of the green pre-TSS sequences (p-value = 2.058e-05).

The manual search for PEP promoter motifs was performed by enforcing 50% similarity (3 out of 6) to all the nucleotides and 60% similarity (2 out of 3) to the overrepresented ones in the published consensus -10 (TAtaaT) and -35 (TTGaca; Liere and Börner, 2007). In case of the -10 box only hexamers with at least one adenine residue were considered. The distance of the -10 element to the TSS was restricted from 3 to 9 nt, since these were the extremes found in already published data on plastid promoters (Sun, et al., 1989; Swiatecka-Hagenbruch, et al., 2007). A -35 element was expected 15 to 21 nt upstream of the -10 box (Harley and Reynolds, 1987). The upstream sequences of white TSS in which no motif was detected by MEME were manually screened for the presence of YRTa motif 2-5 nt upstream the TSS. All sequence logos were generated using WebLogo (http://weblogo.berkeley.edu/; Crooks, et al., 2004).

2.2.9 Secondary structure prediction of the 5' regions of primary transcripts

Minimum free energy (MFE) secondary structures of the 5' regions of the 176 and 244 primary transcripts mapped in green and white plastids, respectively, were predicted using RNAfold (part of the Vienna RNA Package version 1.8.5; Hofacker, et al., 1994; Zuker and Stiegler, 1981; McCaskill, 1990; Hofacker and Stadler, 2006; Bompfunewerer, et al., 2008) using the default parameters and the "-p" flag to retrieve the partition function and base pairing probability matrix. The analysis was performed within the first 50 and 100 nt of the 5' region. (Colored structure plots were generated based on these predictions using the relplot.pl utility of

the Vienna RNA Package) Base on the MFE structures, mountain plot values representing the number of enclosing nucleotides per nucleotide position were calculated using a small Python script (http://www.python.org/). The mountain plot distributions were visualized as box plots using R (http://www.r-project.org/).

2.2.10 IntaRNA prediction of *trans*-encoded ncRNA targets

The -50 to +25 region relative to the start codon of annotated genes in the barely genome (NC_008590) was screend for potential targets of the identified ncRNAs from intergenic regions (Appendix E) using intaRNA, a part of the Freiburg RNA tools (http://rna.informatik.uni-freiburg.de: 8080/IntaRNAHelp.jsp; Smith, et al., 2010), with default settings. The 5' and 3' end termini of the ncRNAs are based on or deduced from the dRNA-seq data, respectively.

2.2.11 Validation of dRNA-seq results by an alternative experimental approach

Several TSS and processing sites (5' ends), as well as 3' ends mapped by dRNA-seq were selected and verified by alternative methods like 5'-RACE, 3'-RACE and cRT-PCR analysis.

2.2.11.1 Verification of 5' ends by 5'-RACE

Plastid 5' ends were mapped by 5'-RACE analysis as previously described (Kühn, et al., 2005). Briefly, 1 µg of total RNA from green and white plastids was treated with tobacco acid pyrophosphatase (TAP; Epicentre) or in buffer alone for 60 min at 37 °C. Following organic extraction (25:24:1 v/v phenol/chloroform/isoamyalcohol), RNA was recovered by overnight precipitation with 3 volumes of ethanol/3 M sodium acetate (pH 5.2). RNA was then ligated to 4pmol of the 5' RNA linker (5R_L; Table 1) in the presence of 10mM ATP, 50U T4 RNA ligase (Epicentre) and 40U RNase Inhibitor (Fermentas) and again purified by organic extraction and precipitation as described above. Reverse transcription (RT) was carried out with SuperScript III reverse transcriptase (Invitrogen) and up to 5 target specific reverse primers (Table 1, Description column, cDNA-primers used for cDNA synthesis), followed by another step of organic extraction and precipitation. The products of the RT reaction were amplified in nested PCR reactions with target (Table 1; Description column; 1 and 2- primers used in 1st PCR and 2nd PCR, respectively) and linker specific primers (5R_1 and 5R_2). PCR products were analyzed on 1.5% TAE gels and products of interest were excised, purified using GeneJet Gel

extraction kit (Fermentas) and ligated into pGEM-T (Promega) or pDrive (QIAGEN) following the manufacturers' instructions. Ligation products were transformed into *E. coli* TOP10 (Invitrogen). Bacterial clones containing the plasmid inserts were subjected to DNA sequencing. The primers used in 5'-RACE analysis are listed in Table 1.

Table 1: Oligonucleotides used for 5'-RACE.

Oligo	Sequence (5'-3')	Target	Description
5R_L	GUGAUCCAACCGACGCGACAAGCUAAUGCAAGANNN		RNA linker
5R_1	TGATCCAACCGACGCGAC		1st PCR linker specific primer
5R_2	ACCGACGCGACAAGCTAATGC		2nd PCR linker specific primer
PZ1	TCCTAGATGAAAACATAGCAGAAAA	T*trnG*-1652; -1659	cDNA, 1st PCR
PZ2	GAAAAACTCTTTGCTTTGGATCT		2nd PCR
PZ3	TGATCTGATGTGCTTTCTTGG	T*trnC*-12426	cDNA, 1st PCR
PZ4	AGTTCAAGTTGCTGAGTTGAGAA		2nd PCR
PZ5	CGCGATACAAAGTTCCTGGT	T*trnT*-2150	cDNA, 1st PCR
PZ6	TCCTGGTAGAGAACTTCTTTGAGTC		2nd PCR
PZ7	TAGTTAACTCTTGCAGTGGAACG	T*trnM*-582	cDNA, 1st PCR
PZ8	TCTTGCAGTGGAACGATAGAGA		2nd PCR
PZ9	GGATCAATTCTAACAAGTCACACACT	T*psbN*-495	cDNA, 1st PCR
PZ10	AAGTCACACACTCATATTCCAGAGA		2nd PCR
Z11	CAATTTGCGTCAAAGGTCCTA	T*trnH*-6209	cDNA, 1st PCR
PZ12	TCATCTTACCGCCTTCCGTA		2nd PCR
PZ13	AGGCCCTGAAAGGAGAAGAA	T*trnI*-1694; -1701	cDNA, 1st PCR
PZ14	GCCAACGGACCTCTGTCTAT		2nd PCR
PZ15	GATAATCGAAATTGAAAAGAACTGTC	T*trnV*-1019	cDNA, 1st PCR
PZ16	TCTTTTCTGTATACTTTCCCCGTTC		2nd PCR
PZ17	TGTACCCTTCTATCCAAATCCAA	T*trnN*-1479	cDNA, 1st PCR
PZ18	CCAAATCCAATTTGCATCG		2nd PCR
PZ19	AGCTCGACTTCTTCCTCTTTTC	T*rpl32*-1830; -1829	cDNA, 1st PCR
PZ20	GCTCTTGCTCAGAGAGATATTAAAAGA		2nd PCR
PZ21	TCTATGGTCTTGGACCGTCA	T*rpl32*-1224	cDNA, 1st PCR
PZ22	TTGGATACTTGATTGACCCACTT		2nd PCR

PZ23	TTTACCAAGGATGAGAATGGATCA	T*trnN*-5275	cDNA, 1st PCR
PZ24	TCTTGGATGGAAATTTCTTTTACCT		2nd PCR
PZ25	CACCGAACCATCCGATGTAA	T*psbA*-80	cDNA, 1st PCR
PZ26	TTTCGCGTCTCTCTAAAATTGC		2nd PCR
PZ27	AATCCCCGCTGCCTCCTT	T*trnE*+1	cDNA, 1st PCR
PZ28	TGCCTCCTTGAAAGAGAGATG		2nd PCR
PZ29	CCAATAGAAGCAAGCCCTACG	T*atpH*-212	cDNA, 1st PCR
PZ30	TTGAAATGTGAATTGTCCGAACT		2nd PCR
PZ31	TCGCCCATTCCTCAAAAGAT	T*psaA*-209	cDNA, 1st PCR
PZ32	GAGACCCGCCAACTGTCTTT		2nd PCR
PZ33	CCTGAAATCGAAAATGCCAAA	T*ndhC*-249	cDNA, 1st PCR
PZ34	GCCTTTGTTTCCTCTGTGCTG		2nd PCR
PZ35	TTATCGTTTCACTTTGTCTCGCTTT	T*ndhC*-329; -336	cDNA, 1st PCR
PZ36	TCACTTTGTCTCGCTTTCTCTAGAATC		2nd PCR
PZ37	GGCGTAAATGAAACTTTAGCAA	T*rbcL*-426; -214	cDNA, 1st PCR
PZ38	TCGTAAAACAAAAAGGATATTCAA		2nd PCR
PZ39	TTGCAATTGCCGGAAATACT	T*psaI*-83;-74; -70	cDNA, 1st PCR
PZ40	AATCCGTCATGGAATAGGTGTCT		2nd PCR
PZ41	AAGACGTATCTTGTTCAAGCCAAT	T*petL*-93; PS*petL*-66	cDNA, 1st PCR
PZ42	GCAGCCAGTAGAAAACCGAAA		2nd PCR
PZ43	CTGATCCCCACGCCTGTATT	T*petG*-39	cDNA, 1st PCR
PZ44	CAGTCACGAATAATCCCGCTAA		2nd PCR
PZ45	CCGCATTGAAAATCCTCCTT	T*psaJ*-251; -247;-198	cDNA, 1st PCR
PZ46	TCCGAAATTGTCCTATCCCTAAAA		2nd PCR
PZ47	AAAAATAAGAAGCCAATTCTGTTCA	T*psbB*-376; -360; -355; -324	cDNA, 1st PCR
PZ48	TTCTATACTCGATAAGTGCCAATATGC		2nd PCR
PZ49	ATCCATTGTACCTAACCCTTCTAGG	T*psbB*-176	cDNA, 1st PCR
PZ50	CTGTGTCAGAAAGCGCGAAT		2nd PCR
PZ51	TGCCCAAACGCGGTATATAAG	T*psbN*-46	cDNA, 1st PCR
PZ52	CCAGATATGGAGATGGCGACT		2nd PCR
PZ53	TCTTTTCCTTTTTGTTTGTTTCGAG	T*ndhB*-275	cDNA, 1st PCR
PZ54	TCGTCCGATTCTCCTTCTATTG		2nd PCR
PZ55	GCACGTATTGTTTGTTGACCA	T*ndhI*-79; -99	cDNA, 1st PCR
PZ56	AACCCAGTTACCATAGGGAACATA		2nd PCR

PZ57	TCCCATGTCTTCCGGCTACT	T*rpoC*-979	cDNA, 1st PCR
PZ58	ATCGATAACGCGCCCTCTT		2nd PCR
PZ59	TCGCCTGTTCTTCCATCAAA	T*rpoC1*-599	cDNA, 1st PCR
PZ60	CCGGGTACTCGGGTTCAAAT		2nd PCR
PZ61	ACACCCAATACGTCGAAACTCA	T*ndhK*-158	cDNA, 1st PCR
PZ62	GTCGAAACTCATTGCCCAAG		2nd PCR
PZ63	AGGCCTTGGACACGTAATGG	T*petL*-3023	cDNA, 1st PCR
PZ64	GTCCCCCTGACCAAAACCTC		2nd PCR
PZ65	GGGGATCTTCTATAATTTCGCACA	T*trnL*-838	cDNA, 1st PCR
PZ66	TTTCTTGGTTTCGTCCAGTCA		2nd PCR
PZ67	GGCAGGCAGATTCACATCTC	T*ndhD*-331	cDNA, 1st PCR
PZ68	CCCACACAATCTTCGGTTCTC		2nd PCR
PZ69	CGAAGTTTTACCATAACAAACATTCC	PS*rps16*-79	cDNA, 1st PCR
PZ70	AACAAACATTCCTCTTTTCATTGC		2nd PCR
PZ71	TTCTCCTGAGGTTGTCGGAAT	PS*ycf3*-62	cDNA, 1st PCR
PZ72	CCATTTACACGGGATCTAGGC		2nd PCR
PZ73	AGACCATTCCAAGGCTCCTTTT	PS*ndhK*-57	cDNA, 1st PCR
PZ74	TTCGCCATGCATAAACTAAACC		2nd PCR
PZ75	TCCCTCCCTACAACTCTTGAA	PS*rbcL*-59	cDNA, 1st PCR
PZ76	CACAATAACAGGGTCTACTCGATG		2nd PCR
PZ77	GACGCTGAGGGCATCCTTTA	PS*rps12*-52	cDNA, 1st PCR
PZ78	TTAAGCGCAGCCGTTTTTCT		2nd PCR
PZ79	CAAGGCAAACCCATGGAAAT	PS*psbB*-63	cDNA, 1st PCR
PZ80	CCATGGAAATACCCCTTTAGCA		2nd PCR
PZ81	CCCGATTGTTTCGATTTTTGA	PS*ndhB*-173	cDNA, 1st PCR
PZ82	CGGAATCCCCATGAATAGGA		2nd PCR

2.2.11.2 Verification of mapped 3' ends by 3-RACE

The abundant chloroplast transcript of *rrn16* (5'-P) was used as a 3' linker in 3'-RACE analysis. 1 µg of total RNA from green and white plastids was treated with 50 U of T4 RNA ligase (Epicentre) in the presence of 1 mM ATP (Epicentre) and 40 U of RNase inhibitor (Fermentas) for 60 min at 37°C. Control reactions without the addition of ligase were set up. RNA was cleaned up by organic extraction (25:24:1 v/v phenol/chloroform/isoamyalcohol) and recovered by overnight precipitation with 3 volumes of ethanol/3M sodium acetate (pH 5.2). Reverse-transcription reactions were carried out with *rrn16* specific primer (3R_1) and SuperScript®III (Invitrogen) according to the manufacturer's protocol, followed by a step of organic extraction and overnight precipitation. Nested PCR reactions were performed with a target primer (Table 2; Description column; 1 and 2 - primers used in 1st and 2nd PCR, respectively) and *rrn16* specific primers (3R_1 and 3R_2). PCR products were analyzed on 1.5% TAE agarose gels. The 3'-RACE products were further handled as described above the 5'-RACE products. The primers used in 3'-RACE analysis are listed in Table 2.

Table 2: Oligonucleotides used for 3'-RACE.

Oligo	Sequence (5'-3')	Target	Description
3R_1	CACCACTTCCCGTTCGACTTG	*rrn16*	cDNA synthesis, 1st PCR rrn16 (linker) specific primer
3R_2	CCGTTCGACTTGCATGTGTTA		2nd PCR rrn16 (linker) specific primer
PZ83	TTTCCTTGAAAAGGGTGCTCA	*rps16*	1st PCR
PZ84	AAGGGTGCTCAACCGACAAG		2nd PCR
PZ85	AATTGCGGAAGCTTGGTTTG	*ycf3*	1st PCR
PZ86	GCGCTTACTCCGGGAAATTA		2nd PCR
PZ87	CTCGGGGGTTCCCTAGTTCT	*rps4*	1st PCR
PZ88	TTCGTGTCCAAGCAGAAGGA		2nd PCR
PZ89	TGATAATCATCCGCGCCTTA	*ndhJ*	1st PCR
PZ90	TTTTCCTAGATGAAACGGAATACCT		2nd PCR
PZ91	GTCAGCAGACGAAGCCAAAG	*clpP*	1st PCR
PZ92	TCCGGAAATGTTTAAGGATTGG		2nd PCR
PZ93	GCCCAGTAGCTACGACCGTTT	*petD*	1st PCR
PZ94	TGGAGCAACACTACCCATTGA		2nd PCR
PZ95	TGGAAATAGCATGGAATAAGGTTTG	*rps7*	1st PCR
PZ96	CCTATTCATGGGGATTCCGTA		2nd PCR
PZ97	GGGTAGAGGCTCTCTTATCCAA	*ndhD*	1st PCR
PZ98	CTCGAATTTTTGAGAACCCTTT		2nd PCR

PZ99	GCTGAAGCCGCTATTGGACT	*ndhE*	1st PCR
PZ10	TTGGCTAATTCGAAATCGTCTG		2nd PCR
PZ101	TCCAAAGCAAAGAGTTTTTCTAATT	T*trnG*-1652; -1659	1st PCR
PZ102	ACTTTTCTGCTATGTTTTCATCTAGGA		2nd PCR
PZ103	AAGGAATTCTTTTGATTCAACCAT	T*trnC*-12426	1st PCR
PZ104	GGATACCAAGAAAGCACATCAGA		2nd PCR
PZ105	GGATCAATTCTAACAAGTCACACACT	T*psbN*-495	1st PCR
PZ106	AAGTCACACACTCATATTCCAGAGA		2nd PCR
PZ107	TTGTCAAAGGGATTGGATTGG	T*trnI*-1694; -1701	1st PCR
PZ108	GGACGTTCCAAAAACGGGTA		2nd PCR
PZ109	ACACAAATTTGCAATTATTCATCTACT	T*trnN*-1479	1st PCR
PZ110	TCATCTACTACTGGAATTTGGATTTTT		2nd PCR
PZ111	AAGCTTCTTTTAATATCTCTCTGAGCA	T*rpl32*-1830; -1829	1st PCR
PZ112	GCTTCTTTTAATATCTCTCTGAGCAAG		2nd PCR
PZ113	TCCTTCATCGTGAGACATATAACCA	T*rpl32*-1224	1st PCR
PZ114	AACCATCACTATAAATAAGAACCAGGA		2nd PCR

2.2.11.3 cRT-PCR

cRT-PCR was used to map simultaneously the 5'-P and 3' ends of *psaC* and was carried out as previously described (Pfalz, et al., 2009). Briefly, tRNA from green plastids was self-ligated and purified as described above for the 3'-RACE protocol. The reverse-transcription reaction was carried out with primer PZ117 (reverse) and SuperScript®III (Invitrogen) according to the manufacturer's protocol, followed by a step of organic extraction and overnight precipitation. Primers PZ117(reverse) and PZ118 (forward) were used in the 1st PCR, while PZ117 and PZ119 (forward) were used in the subsequent nested PCR. The cRT-PCR products were analyzed as described above. The primers used in cRT-PCR analysis are listed in Table 3.

Table 3 Oligonucleotides used for cRT-PCR.

Oligo	Sequence (5'-3')	Target	Description
PZ115	TGGACAAGCTCGTACACATTGA	*psaC*	cDNA synthesis primer; 1st/2nd PCR reverse primer
PZ116	GCCTGAAACAACCCGTAGCA		1st PCR forward primer
PZ117	ACAACCCGTAGCATGGCTCT		2nd PCR forward primer

2.2.11.4 Sequencing of 5'-RACE, 3'-RACE and cRT-PCR products

Bacterial clones carrying plasmids containing 5'-RACE, 3'-RACE and cRT-PCR products were sequenced using the vector specific M13_rev primer (5'-GAGCGGATAACAATTTCACACA GG-3'). Template preparation, cycle sequencing, product purification and analysis were carried out by SMB Services in Molecular Biology GmbH, Rüdersdorf. Briefly, the templates for the sequencing reactions were prepared using TempliPhi500 Amplification kit (GE Healthcare) and the sequencing reaction were set using the ABI PRISM^{TM} Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) according to the manufacturers' instructions. Sequencing and product analysis were done on ABI 3130x automatic DNA Sequencer (Applied Biosystems).

3 RESULTS

3.1 The transcriptome of barley albostrians plastids revealed by dRNA-seq

3.1.1 Experimental setup

In this study, 454 sequencing was used to obtain a comprehensive overview of the plastid transcriptome of the monocot plant barley. One of our main interests was to learn more about the role of PEP and NEP in plastid gene expression. Analysis of NEP-dependent transcription is only feasible in plastids with reduced or no PEP activity (Liere and Börner, 2007). Here, the barley line *albostrians* was used as an experimental system. This mutant line is characterized by green, white and striped leaves with identical genotypes. Plastids in white leaves and white parts of striped leaves are ribosome deficient and lack all plastid-encoded proteins including the core subunits of PEP. Therefore, transcription in these mutant plastids is carried out exclusively by NEP. On the other hand, plastids in green areas are phenotypically identical to wild-type chloroplasts and contain both PEP and NEP (Hess, et al., 1993). The transcriptome of green (G; PEP and NEP present) and white (W; only NEP present) *albostrians* plastids of mature first leaves was analyzed by dRNA-seq (Figure 5). To discriminate between primary (5'-PPP) and processed (5'-P) transcripts, two differential cDNA libraries per plastid type were constructed and compared. G- and W- libraries were generated from untreated total plastid RNA, and thus contained both primary and processed transcripts. The G+ and W+ libraries were created from TEX-treated plastid RNA, and were thus enriched for primary transcripts (Sharma, et al., 2010). After clipping of linkers and poly A-tails, similar number of sequence reads (~ 70 000) ≥ 18 nt were obtained for each library. Using the WU-Blast algorithm, 94.5% (G+) and 79.1% (G-) of the sequence reads from green plastids and 14.3% (W+) and 7.1% (W-) of the reads from white plastids were mapped to the chloroplast genome of *H. vulgare* (NC_008590). The lower percentage of plastid reads in white libraries could be explained by the fact that, due to the limited amount of white mutant leaves and the smaller size of their plastids, both intact and semi-intact plastids were used for total RNA preparation. This resulted in the RNA sample from white tissue being more contaminated by nuclear and mitochondrial transcripts than the RNA sample prepared from green tissue. Nevertheless, the obtained reads from both W libraries were sufficient to provide a good overview picture of the transcriptome of PEP-lacking plastids.

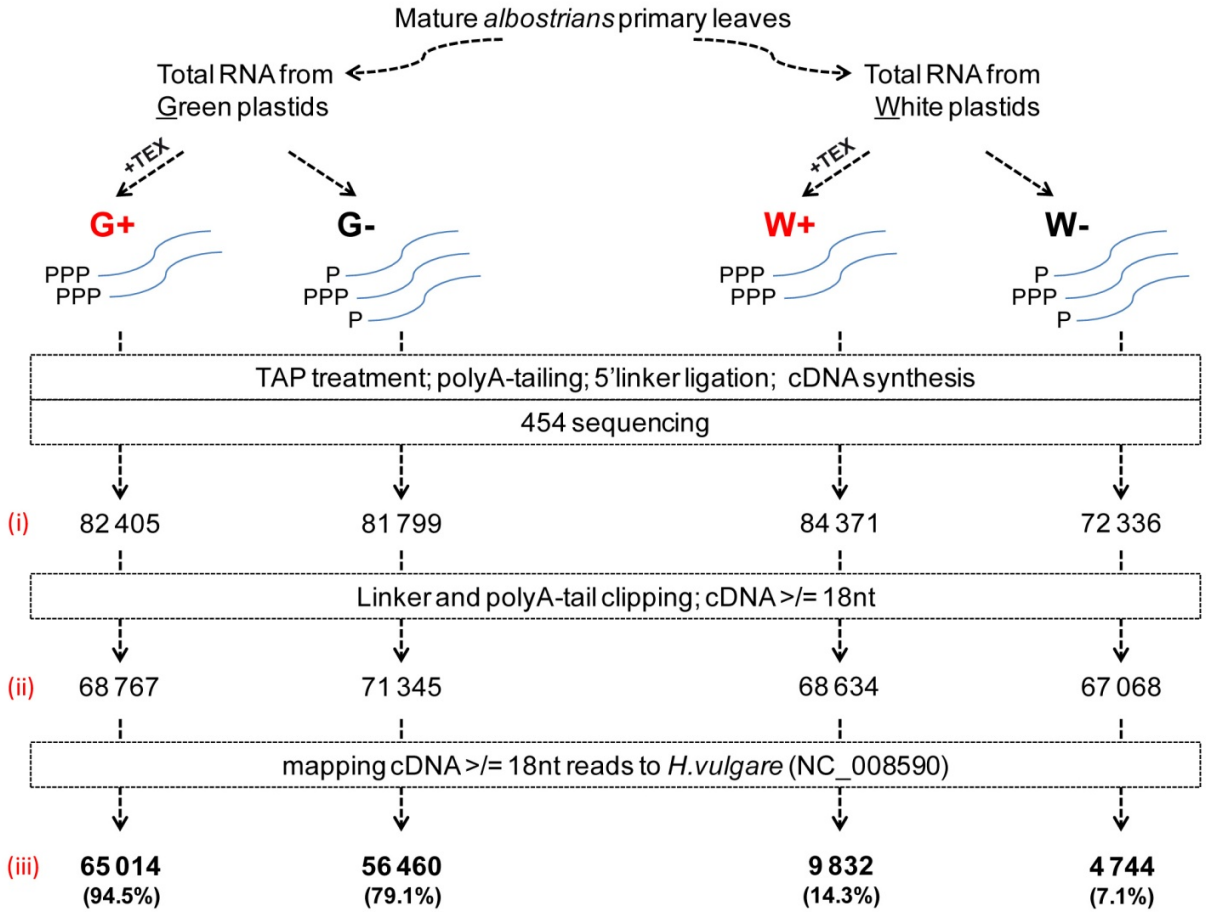


Figure 5: Experimental setup and overview of sequenced and mapped reads. Total RNA from green (G) and white (W) *albostrians* plastids of mature first leaves was used to generate two differential cDNA libraries per plastid type. G- and W- libraries were constructed from TEX untreated RNA which contained both primary (5'-PPP) and processed (5'-P) transcripts. G+ and W+ libraries were generated from RNA treated with TEX, which degrades processed (5'-P) transcripts and thus enriches for primary (5'-PPP) transcripts. RNA was further treated with TAP (tobacco acid pyrophosphates), which converts 5'-PPP to 5'-P (to allow for the subsequent 5' linker ligation), followed by addition of poly(A) tails, 5' linker ligation and reverse transcription. Libraries were sequenced on a Roche 454 FLX sequencer. **(i)** Indicates the number of sequenced reads for each library. After linker and polyA-tail clipping, only cDNA reads longer than/ equal to 18 nt were further considered. **(ii)** A similar number of sequence reads for each library were blasted against the barley chloroplast genome (NC_008590) using the WU-Blast algorithm (http://blast.wustl.edu/). **(iii)** Shows the number of reads successfully mapped to the barley plastome.

For each library, graphs representing the number of mapped sequence reads per nucleotide were calculated and visualized using the Integrated Genome Browser (Figure 6). In order to be able to compare the amount of mapped reads per genome position among the differential library sets, the graphs were further normalized to the number of mapped reads per library. For convenience, cDNA reads belonging to the inverted repeat (IR) regions were mapped only to one

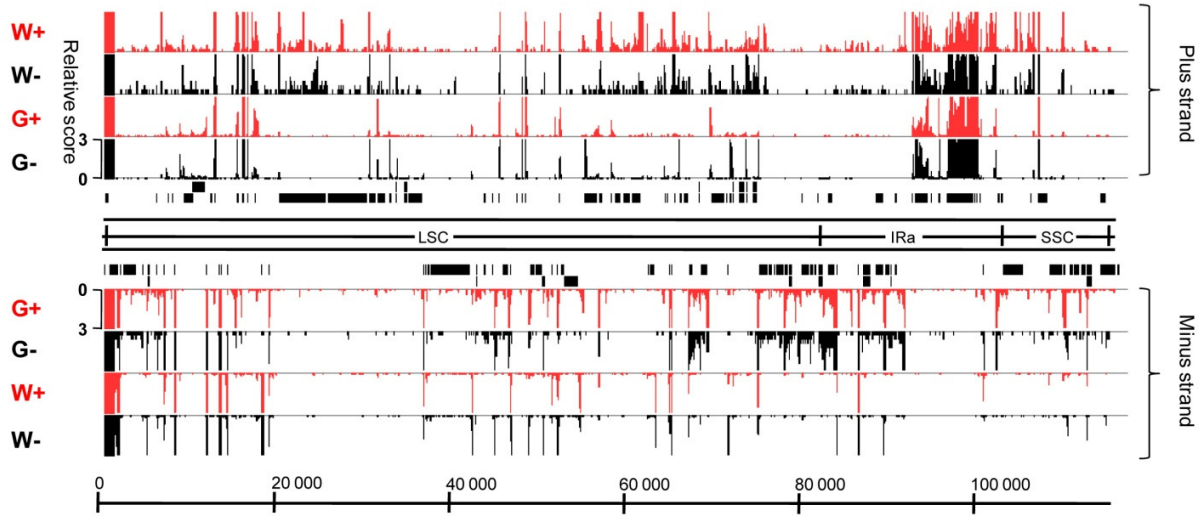


Figure 6: Mapped reads of green (G) and white (W) dRNA-seq libraries. cDNA reads from libraries enriched by TEX treatment (red, (+) libraries) and non-enriched (black, (-) libraries) for primary transcripts were mapped to the barley chloroplast genome (NC_008590). Graphs were normalized to the number of mapped reads per library and visualized using the Integrated genome browser (IGB). The Y-axis indicates per mill (a tenth of a pecentage) mapped reads per genome position. Annotated genes are represented as black boxes. The chloroplast genome of higher plants is divided into four regions: large single copy (LSC), small single copy (SSC) and two inverted repeat (IRa/b) regions. Here, cDNA reads belonging to the IR were mapped only to IRa. Both the plus and the minus are shown.

of them (IRa). cDNAs, covering the entire plastome, were detected in libraries of both green and white plastids. Moreover, a differential cDNA distribution, resulting from TEX activity, was observed in the (+) and (-) libraries.

3.1.2 Discrimination between primary and processed plastid transcripts

dRNA-seq has already proven to successfully discriminate between primary (5'-PPP) and processed (5'-P) bacterial RNAs (Mitschke, et al., 2011; Sharma, et al., 2010). Plastid 5' RNA ends are believed to be identical to 5' RNA ends in bacteria (Stern, et al., 2010), but the ability of the TEX-based method to catalog primary and processed plastid transcripts has so far not been demonstrated. In general, the TEX treatment eliminates 5'-P transcripts and thereby enriches for the ones carrying a 5'-PPP moiety in relative terms. This leads to a characteristic difference in the cDNA distribution in (+) compared to (–) libraries and allows for the mapping of transcription start sites (TSSs; primary transcripts with 5'-PPP) and processing sites (PSs; processed transcripts with 5'P) on a global scale. As expected, the described experimental setup allowed us to distinguish between primary and processed transcripts in plastids, as previously shown for bacteria. The cDNA reads mapped onto *psbB* operon in green plastids provide a proof

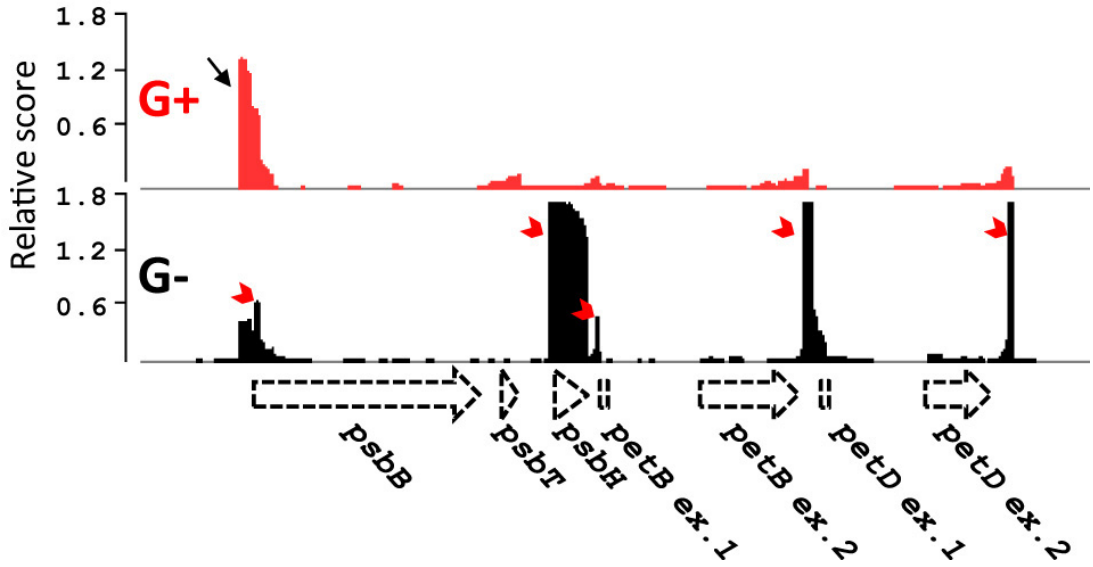


Figure 7: dRNA-seq profile of *psbB* operon in green plastids. cDNA reads of G+/G- dRNA seq libraries mapped onto the *psbB* operon. Genes are depicted by dashed, unfilled arrows/rectangles and triangles below the graph files. The cDNA peaks in G-, which are no longer detectable after the TEX treatment (red arrowheads) represent major processing sites of the *psbB* operon. In the G+ library, the cDNAs are clustered towards the nuclease-protected 5' end of the primary transcript, indicating the TSS of the *psbB* operon (black arrow).

of principle for the unequivocal discrimination of 5'-PPP and 5'-P plastid transcripts by dRNA-seq (Figure 7). This operon comprises five genes transcribed into a large precursor mRNA, which is extensively processed (Bollenbach, et al., 2007). In the G- library (black) several cDNA peaks (indicated by red arrowheads) are visible, which are completely eliminated by the exonuclease treatment and missing in the G+ library (red). The 5' ends of the TEX-eliminated peaks correspond to the previously described major processing sites of *psbB* operon (Barkan, et al., 1994; Chen and Stern, 1991; Felder, et al., 2001; Pfalz, et al., 2009; Westhoff, 1985). In the G+ library, the cDNAs mapped onto *psbB* operon are clustered towards the nuclease-protected 5' end of the primary transcript, clearly indicating the TSS (black arrow) of the operon. The TSS mapped by dRNA-seq (Appendix A) is in agreement with the one previously found in spinach (Westhoff, 1985) and was verified here by 5'-RACE (Appendix B). The TSSs and PSs mapped by dRNA-seq in this work will be further discussed in Chapter 2: The primary transcriptome of *albostrians* plastids and Chapter 3: The processed transcriptome of barley chloroplasts, respectively.

3.2 The primary transcriptome of *albostrians* plastids

3.2.1 Annotation of TSSs revealed by dRNA-seq

In total, 176 and 244 potential TSSs were manually assigned in green and white *albostrians* plastids, respectively (Appendix A). The criteria for annotation of TSSs are summarized in the Materials and Methods Chapter 2.6: Annotation of TSSs and PSs based on dRNA-seq. Surprisingly, in both green and white plastids, more TSSs were identified than one might expect for a genome comprising just 113 genes (Saski, et al., 2007), many of which are clustered in polycistronic transcription units (Kanno and Hirai, 1993). 76% and 91 % of the TSSs mapped in green and white plastids, respectively, were detected at least 2 fold enriched in the (+) versus (-) libraries. Only a small number of TSSs, mapped during the present study by 5'-RACE or described previously, were not detectable as enriched transcripts in the TEX-treated G+ library (Appendix A-1, Not enriched). The initiating nucleotide was a purine in 91% (green) and 84% (white) of the TSSs. In both green and white plastids, A was preferred over G as an initiating nucleotide, more pronounced in the latter.

3.2.2 Validation of the TSSs revealed by dRNA-seq

In order to clarify the question if dRNA-seq is a suitable method to reliably identifies plastid transcription initiation sites, the TSSs mapped in this study (Appendix A) were subjected to further analysis using alternative means: (i) comparison with available data in the literature; (ii) computational analysis; and (iii) 5'-RACE analysis of selected candidates.

In total, 11 out of the 12 TSSs, which were previously mapped in barley using alternative methods, could be detected and defined with high accuracy by dRNA-seq (up to +/- 2 nt difference; Table 4). A possible reason for the undetected T*rbcL*-316 (TSSs are abbreviated with a T and named after the downstream located gene and the number of nucleotides between the TSS and the start codon of the corresponding ORF or the mature 5' end of the corresponding tRNA/rRNA) could be that the abundance of this primary transcript in mature chloroplasts is below the level of detection by dRNA-seq in this study.

Table 4 Comparison of TSSs determined by dRNA-seq with previously mapped primary ends

TSS[1]	Strand	Previously mapped genomic position	Genomic position based on dRNA-seq	Dif.[3] (nt)	Reference[4]
T*psbA*-80	-	1760	1760	0	Boyer and Mullet, 1988
T*psbK*-171	+	7096	7096	0	Sexton, et al., 1990; Sexton, et al., 1990
T*psbD*-711	+	8448	8448	0	Sexton, et al., 1990; Sexton, et al., 1990
T*psbD*-557	+	8602	8602	0	Sexton, et al., 1990; Sexton, et al., 1990
T*psbC*-194	+	9972	9974	2	Sexton, et al., 1990; Sexton, et al., 1990
T*psaA*-209	-	42091	42089	2	Berends, et al., 1987; Swiatecka-Hagenbruch, et al., 2007 (*Arabidopsis*)
T*rbcL*-316[2]	+	54623	n.d.	-	Poulsen, 1984
T*clpP*-132	-	69033	69032	1	Hübschmann and Börner, 1998
T*rpl23*-71	-	83582	83580	2	Hübschmann and Börner, 1998
T*rpoB*-147	+	19940	19940	0	Liere and Börner, 2007; Silhavy and Maliga, 1998 (Maize)
T*atpB*-593	-	54749	54749	0	Liere and Börner, 2007; Silhavy and Maliga, 1998 (Maize)
T*rrn16*-116	+	92567	92569	2	Hübschmann and Börner, 1998

[1]The TSSs are marked with a T and named after the downstream located gene and the number of nt between the primary 5' end mapped in this study and the start codon of the ORF (e.g., T*psbA*-80) or the 5' end of the mature rRNA (e.g. T*rrn16*-116).
[2]T*rbcL*-316 is named after the number of nt between the primary 5' end mapped in the reference, since no primary 5' end was mapped in this study.
[3]The difference (in nucleotides) between the previously mapped genomic position of a TSS and the one determined here is calculated.
[4]The references of the previously determined TSSs are provided.

Theoretically, a processed 5' end might be protected from TEX digestion by stable structures resulting from intramolecular base pairing (Sharma, et al., 2010). In such a case, a 5' end could be misannotated as a putative TSS instead of a PS. Therefore, the probability of stable structure formation in the first 50 and 100 nucleotides of alpotential primary transcripts mapped in this work was determined computationally (for details see Chapter 2.2.9.). On average, the number of

nucleotides potentially participating in base pair interactions was very low in the region immediately downstream of the mapped TSSs (Figure 8). That is, stable stem loops are unlikely to be formed near the 5'end of most analyzed transcripts and thus most likely do not act as a barrier to digestion by TEX. Moreover, 52 TSSs, among them examples with predicted relatively stable stem-loops at their 5' end, were selected and 40 TSSs (77%) were successfully verified by

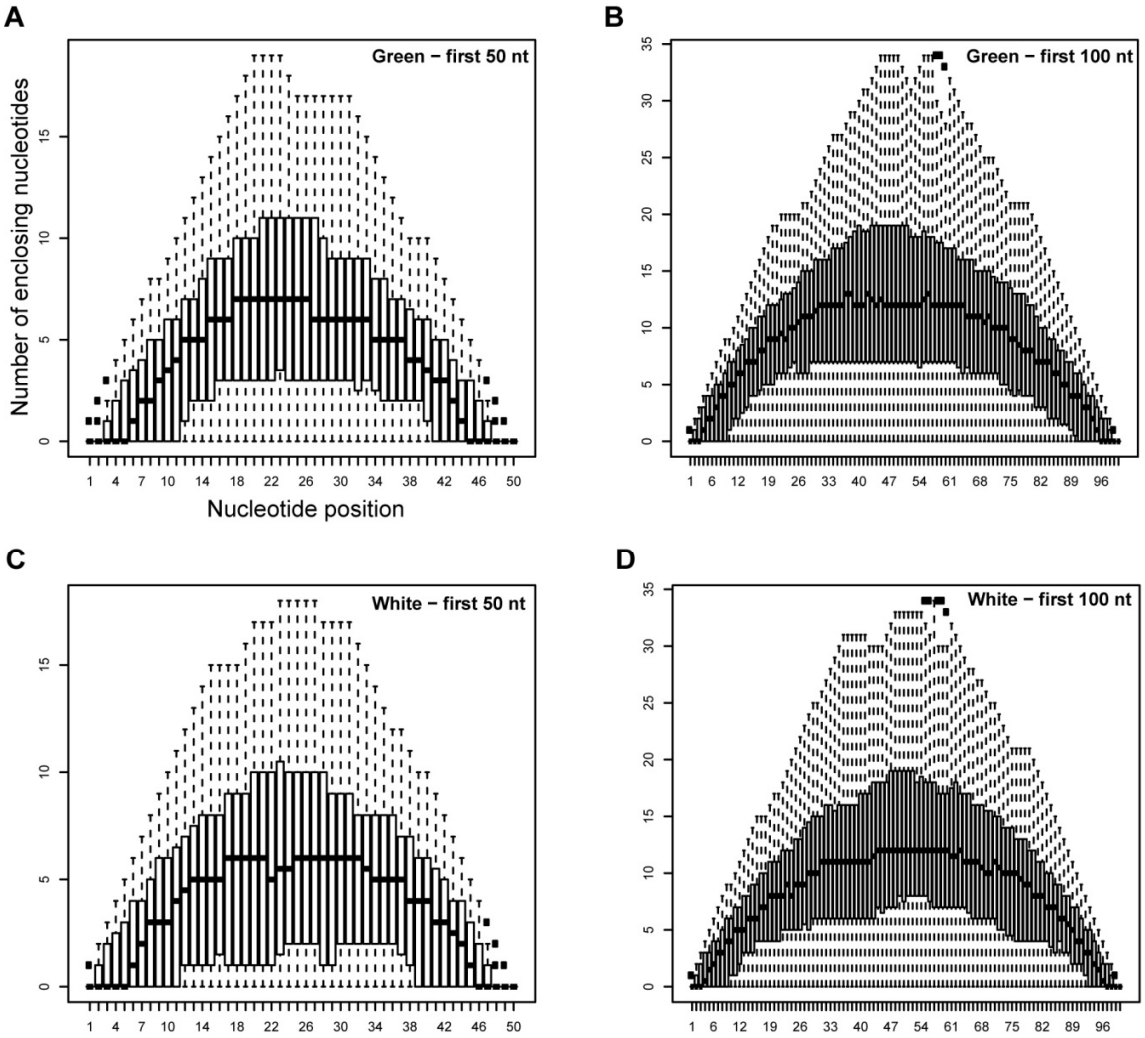


Figure 8: Prediction of stable structure formation at the 5' ends of primary transcripts. Mountain plot value distributions representing the number of enclosing nucleotides per nucleotide position within the first 50/100 nt of all primary transcripts in green/white plastids. The mountain plot values were calculated based on the minimum free energy structures predicted of the analyzed sequences. **(A)** and **(B)** Mountain plot value distribution for the first 50 and 100 nt, respectively, of all primary transcripts in green plastids. **(C)** and **(D)** Mountain plot value distribution for the first 50 and 100 nt, respectively, of all primary transcripts in white plastids.

5'-RACE analysis, as major 5'-PPP ends in independent RNA preparations from green and white plastids. Eight TSSs were detected as minor 5'-PPP ends by this analysis, i.e., found in less than 20% of the sequenced clones of the corresponding TAP+ reactions or detected less abundant than other 5'-PPP mapped in the same experiment. 5'-RACE could not detect three TSSs, and in the case of one the method could not distinquish if this 5' end is primary or processed (Appendix B). Taken together, these results indicate that the TEX-based RNA-seq approach used in this work reliably identifies TSSs and discriminates them from PSs in RNA preparations from plastids.

3.2.3 Classification of TSSs revealed by dRNA-seq

The assigned 176 and 244 TSSs in green and white *albostrians* plastids, respectively, were grouped into 4 categories based on their location with respect to annotated genes in the barley plastome (Figure 9A). gTSSs (gene TSSs) were located within the 750 nt region upstream of annotated mRNA start codons or of 5' mature ends of rRNA or tRNA genes. iTSSs (internal TSSs) were mapped within annotated genes and gave rise to sense transcripts. aTSSs (antisense TSSs) were found on the opposite strand within or up to 150 bp upstream or downstream of annotated genes, and thus gave rise to antisense transcripts. oTSSs (orphan TSSs) could not be assigned to any of the above categories and are mapped to intergenic regions. Several TSSs could be assigned to more than one category.

The aforementioned categories were defined as previously described (Sharma, et al., 2010) with the following modifications. The distance of gTSSs to the annotated genes was increased from 500 to 750 nt since chloroplast genes can be transcribed from promoters located far upstream by both PEP and NEP (Liere and Börner, 2007). In barley the most distant experimentally verified gene initiation site is T*psbD*-711 (Sexton, et al., 1990; Sexton, et al., 1990), which was also mapped here as a TSS in green plastids by dRNA-seq (Appendix A-1). However, the cDNA reads corresponding to T*psbD*-711 did not reach the *psbD* ORF, most likely due to the 454 FLX sequencer maximum read length of 400 bases. gTSSs with cDNA reads that did not reach into the corresponding downstream gene are referred to as "disconnected" (see *Comments* section of Appendix A). Still the possibility remains that some of the distant and disconnected gTSSs might be involved in ncRNA rather than mRNA synthesis. Moreover, it is also possible that some genes are transcribed by promoters located beyond the 750 nt upstream region considered in this analysis (Vera, et al., 1996). In such cases their TSSs will be

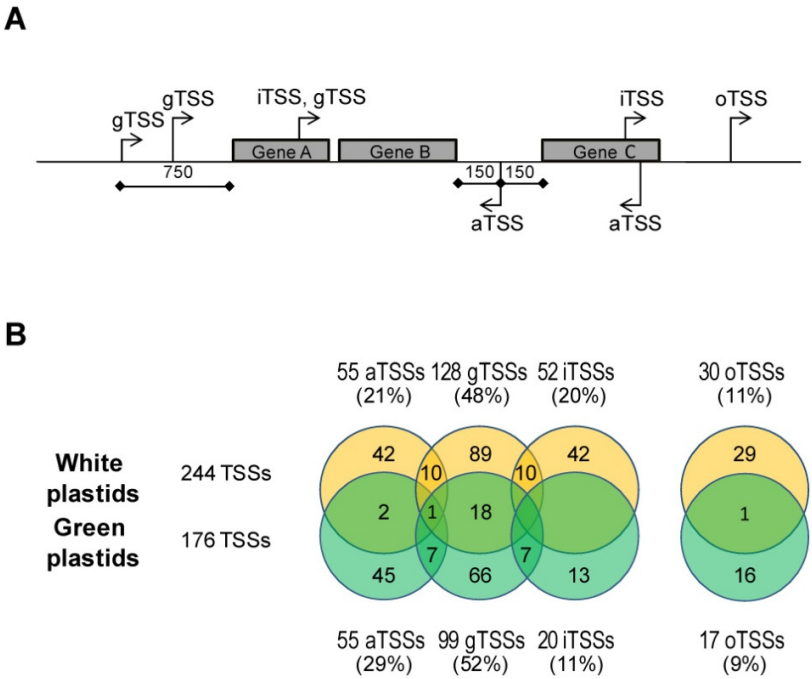


Figure 9: Classification and category assignment of TSSs based on dRNA-seq. (A) Schematic representation of the annotation and category assignment of TSSs based on their genomic location in the barley plastome. In certain cases, a TSS can be assigned to more than one category, e.g., iTSS and gTSS. **(B)** Distribution and overlap among TSS categories in green and white plastids. Two-hundred-and-forty-four and 176 TSSs were mapped in white and green plastids, respectively. TSSs were further grouped into four categories. The number and percentage (in brackets) of TSSs assigned to each category is given. Twenty-one TSSs in white and 15 TSSs in green could be assigned to more than one category. Twenty-two TSSs were found in libraries of both green and white plastids.

misannotated as TSSs involved in ncRNA synthesis. The 5'- and 3'-UTRs (untranslated regions of mRNAs) detected in this study were on average longer than 100 nt (see Chapter 3.2.6.3. and Appendix I/J, respectively). Therefore, the distance of an aTSS to the upstream/downstream annotated gene on the opposite strand was increased to 150 nt. An aTSS that is mapped more than 150 nt downstream of an annotation but is represented by cDNAs that overlap with the 3' UTR of the gene is still considered as an aTSS.

In both green and white plastids TSSs belonging to all four categories were detected (Figure 9B). The majority of TSSs in both plastid types (52% and 48% of all TSSs mapped in green and white, respectively) were gTSSs, *i.e.* involved in the expression of annotated genes. The second most abundant category was aTSSs, with 29% of the TSSs mapped in green and 21% in white, *i.e.* a high degree of antisense transcription was detected in barley plastids. Surprisingly, numerous TSSs were detected within annotated genes in both green and white libraries, making iTSSs the third most abundant class of TSSs. The least populated category was oTSSs, most

likely due the fact that the plastome of higher plants is densely packed with coding sequences (Sugiura, 1992). Twenty-one TSSs in white and 15 TSSs in green could be assigned to more than one category, *e.g.* a TSS could be both an iTSS and a gTSS (Figure 9A).

3.2.4 Comparison of TSSs mapped in green and white plastids

The comparison of TSSs mapped in green (both PEP and NEP present) and white (only NEP present) *albostrians* plastids led to interesting observations. Strikingly, only 22 TSSs, the majority of which were gTSSs, were found to be identical in both plastid types (Figure 9B; Appendix C). Since NEP activity was reported to be higher in white compared to green leaves (Emanuel, et al., 2004), chloroplasts of green leaves are not expected to utilize NEP promoters that are not found in white plastids. Therefore, at least 154 of the 176 TSSs (88%) in green barley chloroplasts should have originated from PEP activity. That is, PEP is by far the dominating RNA polymerase in chloroplasts of mature barley leaves. Moreover, 222 (91%) of the TSSs mapped in white plastids could not be found in green plastids. Therefore, the absence of PEP leads to the activation of numerous NEP promoters, which cannot be detected in mature chloroplasts.

3.2.5 Promoter sequence analysis of TSSs mapped in green and white plastids

The -1 to -25 and -26 to -50 nt upstream regions of all TSSs mapped in green and white plastids were screened separately for potential promoter motifs with length of 3 to 9 nt using MEME (Multiple Expectation-Maximization for Motif Elicitation), a tool for discovering motifs in a group of related sequences (Bailey, et al., 2009). Two 8 nt-long motifs were discovered in the -1 to -25 nt and -26 to -50 nt upstream region of only 44 (25%) and 20 (11%) of the TSSs, respectively, in green plastids (Figure 10A). These two motifs were found to be significantly overrepresented (p-value = 2.434e-05 and 0.008552, respectively) in pre-TSS stretches of green plastids in comparison to the ones of white plastids and show high similarity to the previously described -10 and -35 PEP promoter consensus hexamers (Liere and Börner, 2007). The MEME output was in accordance with the observation that PEP is the dominating polymerase in green *albostrians* plastids (responsible for 88% of the TSSs detected in this study). However, it was surprising that MEME discovered PEP motifs in such a small percentage of the upstream sequences of the TSSs mapped in green. A possibility for the poor detection of promoter

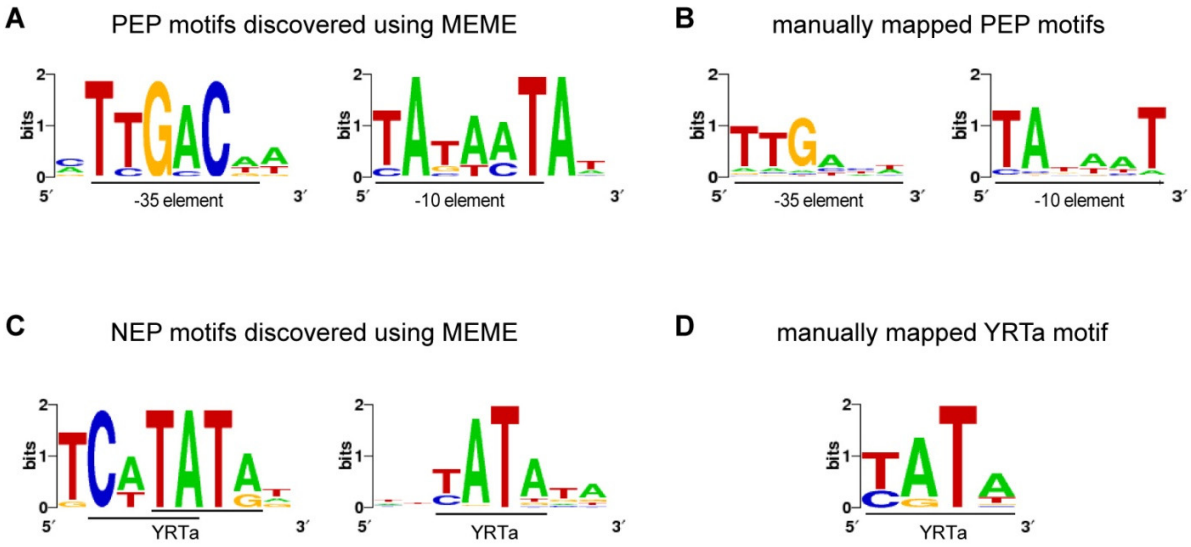


Figure 10: Sequence logos of promoter motifs detected in green and white plastids. Logos were visualized using WebLogo (http://weblogo.berkeley.edu/). **(A)** In green plastids, MEME analysis discovered a -10 (right) and -35 (left) PEP consensus element upstream of 44 and 20 TSSs, respectively. The motifs were found to be significantly enriched in green pre-TSS sequences. **(B)** A manual search for the PEP promoter elements detected the -10 box (right) in 156 TSSs and the -35 box (left) in 109 of the TSSs with a mapped -10 element. **(C)** Two versions of the YRTa motif were discovered by MEME in white plastids. A TCaTATat motif (left) was found upstream of 22 of the white TSSs and YATata (right) upstream of 151 (62%) TSSs. **(D)** YRTa motif was manually mapped in 73% of the TSSs in white plastids.

consensus sequences could be that the -10 and -35 elements are more variable than in bacteria, and therefore hard to identify using default cut-off values of the MEME tool. In order to investigate this assumption, a manual search with relaxed parameters was performed to further analyze the occurrence of these motifs (for details see Chapter 2.2.8.). Indeed, a -10 element (TAtaaT; upper-case letters depict overrepresented nt >1 bit) 3-9 nt upstream of the transcription start point in 156 TSSs (89% of the TSSs) mapped in green plastids could be detected. The -35 region was again found to be less conserved than the -10 box. A ttGact motif 15-21nt upstream of the -10 element was mapped in 70% (109/156) of the TSSs (Figure 10B).

A MEME search of sequences from white plastids (only NEP present) revealed an 8-nt long motif in the -1 to -25 upstream region of 22 TSSs (Figure 10C, left). The motif was found significantly overrepresented (p-value = 0.009847) in white compared to green pre-TSS stretches and resembled an extended version of the YRTa motif, the most frequently observed NEP promoter motif (Liere and Börner, 2007). An additional search limited only to the first 10 nucleotides upstream of the TSSs revealed a second motif significantly more predominant (p-

value = 2.058e-05) in white pre-TSS regions. The motif represents a YRTa core flanked by AT-rich sequences and was found in 151 (62%) of the white TSSs (Figure 10C, right). Unfortunately, MEME did not detect any additional motifs in the upstream regions or even when the analysis was extended to the -50 to +25 regions around the 244 TSSs in white plastids. Moreover, a manual search for the YRTa motif increased its detection to only 73%, further suggesting a low conservation of NEP promoter elements (Figure 10D).

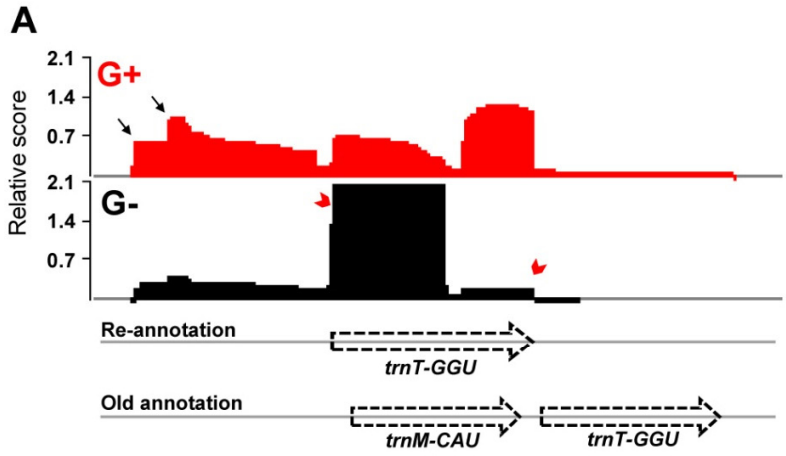
3.2.6 The primary transcriptome of annotated genes

3.2.6.1 Re-annotation in the barley plastome based on dRNA-seq

Two *trn* genes are annotated on the plus strand in the region from 15100 to15400 on the barley chloroplast genome available in the nucleotide database of NCBI under the accession number NC_008590: *trnM-CAU* (from 15209 to 15267) and *trnT-GGU* (from 15275 to 15338). However, cDNAs corresponding only to *trnT-GGU*, mapping to position 15203 to 15274 were detected in dRNA-seq libraries of both green and white plastids (Figure 11A; only cDNAs from green plastids are shown). The absence of *trnM-CAU* and the new annotation of *trnT-GGU* were supported by tRNAscan-SE program (http://lowelab.ucsc.edu/tRNAscan-SE/; Lowe and Eddy, 1997), a software designed to search for tRNA genes in genomic sequences (Figure 11B). Moreover, based on dRNA-seq data and supported by tRNAscan-SE, *trnK-UUU* coordinates were also reannotated: exon 1, 4425 to 4460; exon 2, 1910 to 1945 (data not shown).

3.2.6.2 Gene count and operon annotation of the barley chloroplast genome

Based on the published sequence and our re-annotation (see above), there are 113 unique genes on the barley chloroplast genome (NC_008590), 78 of which are protein-coding, and 37 are coding for tRNAs or rRNAs. In the current calculation: (i) the genes from IR regions were counted only once; (2) the missing *trnM-CAU* (see Chapter 3.2.6.1.) was not included in the count; (iii) maturation of *rps12* mRNA involves *trans*-splicing of *5'-rps12* and *3'-rps12* (Hildebrand, et al., 1988) and the respective genes were therefore considered as two separate genes. Gene content and gene order of the barley chloroplast genome was found to be identical to the one of rice (*Oryza sativa*; Saski, et al., 2007). According to experimental data 17 polycistronic and 22 monocistronic transcripts were detected in rice (Kanno and Hirai, 1993).



B

```
Your-seq.NC_008590 (15203-15274)      Length: 72 bp
Type: Thr      Anticodon: GGU at 33-35 (15235-15237) Score: 67.30
         *    |    *    |    *    |    *    |    *    |    *    |    *    |
Seq: GCCCTTTTAACTCAGTGGTAGAGTAATGCCATGGTAAGGCATAAGtCATCGGTTCAAATCCGATAAAGGGCT
Str: >>>>>>>..>>>>.......<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<.
```

Figure 11: Re-annotation based on dRNA-seq. (A) cDNA reads of green (G+/-) dRNA-seq libraries mapped onto the region from 15100 to 15400 on the barely plastome. *trnM-CAU* and *trnT-GGU* were previously annotated in this region (see old annotation). cDNAs corresponding to only *trnT*-GGU (see re-annotation) were detected in dRNA-seq libraries. The TSSs of *trnT* are indicated by black arrows. The mature 5' and 3' end of the tRNA are marked by red arrowheads. **(B)** tRNAscan-SE output. tRNAscan-SE supports the re-annotation of *trnT* from 15203 to 15274.

Here, the same gene organization was accepted for barley plastids but with the following modifications: (i) *trnE-Y* was regarded as an operon since the corresponding polycistronic transcript was detected in this work; (ii) the *rpoC1-C2* bicistronic transcript was expanded to *rpoB-rpoC1-rpoC2* (Hudson, et al., 1988); (iii) the *clpP*-5'*rps12* operon was corrected to *clpP*-5'*rps12-rpl20* (Hübschmann, et al., 1996); (iii) *trnK exon 1-matK-trnK exon 2* was regarded as a polycistronic transcript; (iv) *psbM*, *psbN,* and *ndhF* were considered to be monocistronic transcripts, as previously shown (Casano, et al., 2001; Kawaguchi, et al., 1992); (v) *trnG-trnfM* was considered to be a polycistronic transcript (Oliver and Poulsen, 1984); (vi) *psaI, rpl23* (HvsvCp031), *trnV-GAC* and *trnS-UGA* failed to be detected in rice as parts of any of the nearby transcription units and were therefore assumed to be transcribed monocistronically. Taken together, the 89 barley plastid genes are proposed to be grouped into 20 operons, while 24 genes are transcribed as monocistronic RNAs (Figure 12).

Figure 12: Operon and TSS map of the barley chloroplast genome. The outer circle depicts the gene organization of the barley chloroplast genome (NC_008590). The graphical representation was created using OrganellarGenomeDRAW; Lohse, et al., 2007) and further modified. Genes at the inside and outside of the circle are transcribed clockwise and counter clockwise, respectively. Assigned operons (see Chapter 3.2.6.2.) are marked by arrows. Genes are color coded based on the detection of their TSSs in the corresponding plastid type: green - genes for which TSSs were detected solely in green plastids; yellow - genes for which TSSs were detected solely in white plastids; red- genes for which TSSs were detected in both plastid types; and grey - genes for which TSSs were not detected in our analysis. The inner circle of the figure depicts the genomic position of all mapped TSSs as follows: green - TSSs mapped in G library; orange - TSSs mapped in W library and red - TSSs identical between G and W. cDNA reads mapped to the inverted repeat (IR) are shown only within IRa. The image was generated using CGView (Circular Genome Viewer;Stothard and Wishart, 2005).

3.2.6.3 Overview of gTSSs in green and white plastids

As already mentioned, the majority of TSSs in both plastid types detected in this study were gTSSs, *i.e.* involved in the expression of annotated genes. In total, 99 gTSSs were assigned for 64 genes in green plastids, wheras 128 gTSSs were found for 70 genes in white plastids. Taking into account the operon organization of the barley plastome, TSSs upstream of 15 and 14 operons and of 19 and 20 monocistronically transcribed genes were found in green and white plastids, respectively (Figure 12, Appendix D). Still the possibility remains that the monocistronically transcribed genes and operons without visible TSSs are expressed from promoters located beyond the 750 nt upstream region considered in this analysis. For example, the only promoters (a PEP and a NEP promoter) for the tobacco *rpl32* gene are located approximately 1100 nt upstream of the start codon (Vera, et al., 1996). In this range, TSSs for the *trnF-GAA* (TtrnF-938) and *trnI-CAU* operon (TtrnI-1090) were found in green plastids and for the *trnN-GUU* (TtrnN-775), *trnS-UGA* (TtrnS-798), *trnT-rps4-ycf3* (TtrnT-1115) and *psbE-F-L-J* operons (T*psbE*-1075) in white plastids. It can be speculated that these distant TSSs could drive the expression of the corresponding genes and operons. Alternatively, 5'-P transcripts were detected for the majority of the monocistronically transcribed genes and operons without visible gTSSs in this analysis, suggesting either fast mRNA maturation or very low abundance of the primary transcripts in these cases. In the samples of green plastids, was neither a gTSS nor a processed transcript detectable for *rpl32*, *ndhH* and *ndhF* among the evaluated ~70.000 sequences per library. These genes may be transcribed either at an undetectable low level or not at all. In the case of white libraries, no transcripts of *trnL-UAA*, *ndhH*, *ccsA* and *psbE-F-L-J* were identified. A possible explanation for the above listed undetected primary transcripts in white plastids is that these genes may simply have no NEP promoter. Still, the distant TSS T*psbE*-1075 could be involved in the transcription of the *psbE* operons. Surprisingly, 44 gTSSs in green and 58 in white leaves were found within operons and thus, may function in uncoupling those gene clusters into smaller polycistronic and/or monocistronic transcription units. Transcription from multiple promoters is a common feature of plastid transcription (Liere and Börner, 2007). Here, between two and six TSSs, with two being the most frequent, were found for 25 and 35 genes in green and white plastids, respectively (Appendix D). In general, bacteria have a relatively short 5' untranslated region (5' UTR ≤ 40 nt; Mitschke, et al., 2011; Passalacqua, et al., 2009; Sharma,

et al., 2010). In contrast to that, the average length of the plastid 5'UTR detected in this study was much longer: 235 nt and 232 nt in green and white *albostrians* plastid, respectively.

3.2.6.4 Identical gTSSs in green and white plastids

A major goal of this study was to elucidate the division of labor between PEP and NEP in plastid transcription. The comparison of the TSSs mapped in green (both PEP and NEP present) and white *albostrians* plastids (only NEP present) could reveal NEP primary transcripts in mature chloroplasts. 80 gTSSs were found in RNA libraries of green but not white leaf material, therefore representing primary PEP transcripts. 19 gTSSs were found in both green and white plastids (Appendix C). Five of them were TSSs of *trn* genes mapped at position +1 relative to gene start. These TSSs won't be further discussed since their authenticity is still a matter of debate (see Chapter 3.2.6.6.). The initial hypothesis was that the remaining 14 identical gTSSs should most likely represent primary NEP transcripts in green plastids. Indeed, among them were T*rpoB*-147, T*rpl23*-71 and T*rps15*-228, which were previously reported to be NEP-dependent TSSs in green plastids (Hübschmann and Börner, 1998; Liere and Maliga, 1999; Swiatecka-Hagenbruch, et al., 2007). These TSSs were found overrepresented in the W+ compared to the G+ library. The dRNA-seq profile of *rpl23* is shown in Figure 13A. However, identical gTSSs were found in both green and white libraries in the case of the photosynthesis genes *psbA, psaA* and *psbN,* which were shown to be transcribed from PEP promoters in chloroplasts (Swiatecka-Hagenbruch, et al., 2007; Zghidi, et al., 2006). As an example, the dRNA-seq data for *psbA* is presented in Figure 13B. Moreover, T*psbA*-80 and T*psbN*-46 were verified by 5'-RACE in independent RNA preparations from both green and white plastids (Appendix B). The identification of these TSS in white samples cannot be explained by a contamination with green chloroplasts, since the PEP-dependent TSSs of other abundant RNAs (e.g. *trnG-UCC*, *psbE*, *petN*) were not detected in samples from white tissue. A more plausible reason for the occurrence of these identical primary transcripts in both plastid types might be that the promoter regions of T*psbA*-80, T*psaA*-209, and T*psbN*-43 represent overlapping motifs recognized by both PEP and NEP. Indeed, promoter motifs for both polymerases were found upstream of T*psbA*-80 and T*psbN*-43 (Appendix C). Therefore, it is obviously possible that PEP and NEP can synthesize transcripts with identical 5' ends. Consequently, it could not be precisely determined if PEP, NEP or both polymerases generate the rest of the identical TSSs in green barley chloroplasts,

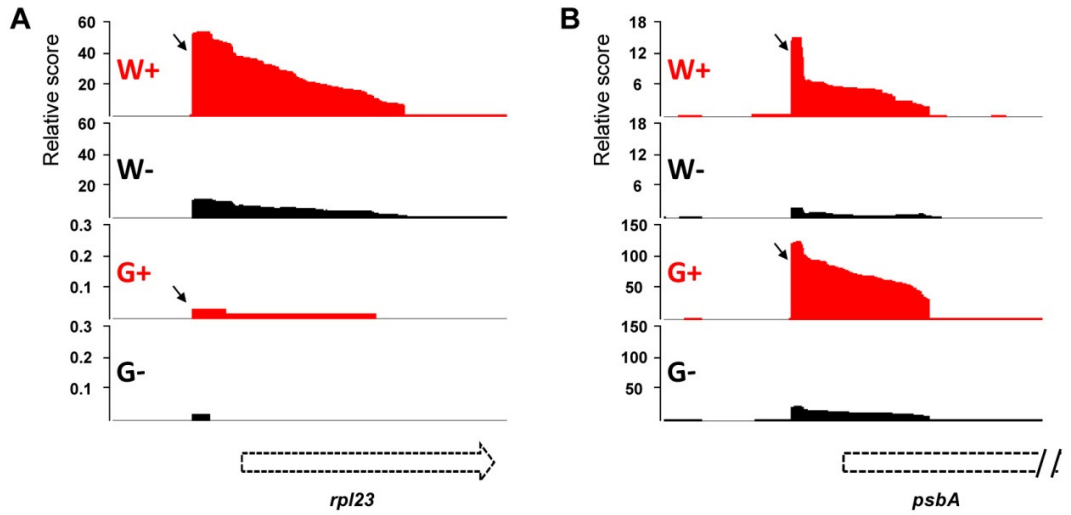


Figure 13: Identical gTSSs in green and white dRNA-seq libraries. (A) cDNA reads of green (G+/-) and white (W+/-) dRNA-seq libraries mapped onto *rpl23*. The same TSS of *rpl23* (T*rpl23*-71; marked with black arrows) was detected in both green and white plastids, significantly overrepresented in the latter. **(B)** cDNA reads of green (G+/-) and white (W+/-) dRNA-seq libraries mapped onto *psbA*. The same TSS of *psbA* (T*psbA*-80; marked with black arrows) was detected in both green and white plastids, significantly overrepresented in the former.

among them the gTSSs T*trnK*-239, T*psbM*-348, T*rps*2-152, T*ndhC*-336, T*trnP*-21, T*rps8*-142; T*ndhB*-275, T*ndhI*-99. Based on the above listed aspects, several plastid genes were grouped in the following categories: (i) genes clearly transcribed *via* a NEP promoter in chloroplasts - *rpoB*, *rpl23*, and *rps15*; (ii) genes potentially transcribed *via* a NEP promoter in chloroplasts - *trnK*, *psbM*, *rps2*, *ndhC*, *trnP*, *rps8 ndhB* and *ndhI*. For all of the above mentioned genes, except *rpoB*, *rpl23* and *ndhB*, at least one additional PEP-dependent TSS was detected in green plastids. Therefore, this analysis revealed *rpoB* and *rpl23* as the only genes that are, without any doubt, transcribed exclusively by NEP in mature barley chloropalsts.

3.2.6.5 Genes with NEP promoters in white *albostrians* plastids

As already mentioned, the in-depth mapping of NEP promoters is only feasible in experimental systems with reduced or no PEP activity. The reason for this is that in wild-type chloroplasts, NEP activity is scarce, while PEP is functioning as the main polymerase (Liere and Börner, 2007). Indeed, this phenomenon was also observed here (see Chapter 3.2.4.). In this study, the dRNA-seq analysis of the transcription of PEP-deficient white *albostrians* plastids allowed for the identification of numerous NEP promoters that could not be detected in mature chloroplasts.

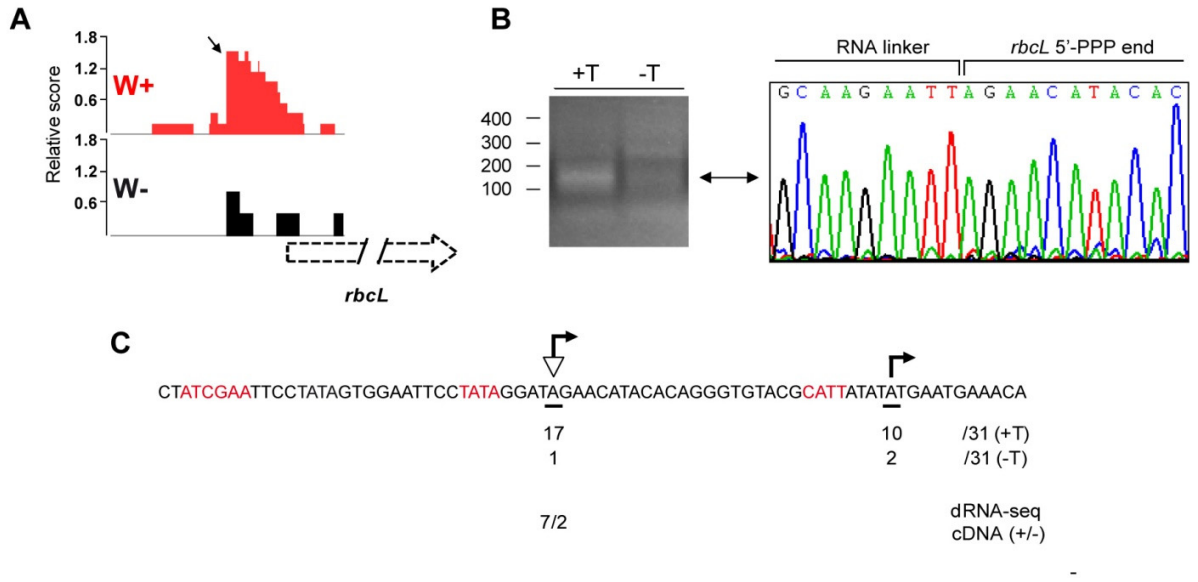


Figure 14: *rbcL* is transcribed from a NEP promoter in white plastids. (A) cDNA reads of white (W+/-) dRNA-seq libraries mapped onto *rbcL*. The TSS of *rbcL* (T*rbcL*-214) is indicated by a black arrow. **(B)** Left: 5'RACE T+ (TAP+) and T- (TAP-) reactions separated on a 1.5%TAE gel. The product of expected size is marked by an arrow. Right: A Chromatogram displaying sequence at the ligation site of a cloned 5'-RACE product from the T+ reaction. The 3' end of the linker and the 5'-PPP end of the *rbcL* are shown. **(C)** Promoter region of *rbcL* used in white plastids. The TSS identified by dRNA-seq and 5'-RACE are marked by a triangle and arrows, respectively. The sequence motifs shown in red represent putative NEP promoters. The number of clones supporting each TSS in T+ and T- reactions, as well as the total number of sequenced clones (number after the slash) is given. The number of cDNAs in (+/-) dRNA-seq is also listed.

Data from previous experiments suggested that several genes like *atpB*, *rrn16* and *rps4* have both PEP and NEP promoters (Swiatecka-Hagenbruch, et al., 2007). The dRNA-seq analysis of *albostrians* plastids revealed that it is a common feature of plastid genes to have promoters for both polymerases (Figure 12, Appendix D). Only for 11 (*trnL-UAA*, *trnM-CAU*, *trnN-GUU*, *psbE-F-L-J*, *trnT*, *petN*, *trnS-UGA*, and *trnQ-UGG*) out of the 64 genes with primary transcripts in green plastids were no TSSs identified in the RNA of white plastids. Those genes may indeed have no NEP promoter. However, it cannot be ruled out that NEP-dependent primary transcripts of these genes exist below the detection level of this study. Moreover, the distant TSSs T*trnN*-775 and T*psbE*-1075 could be responsible for the expression of the corresponding ORFs rather than participate in ncRNAs synthesis (see Chapter 3.2.6.3.). Genes coding for subunits of several photosynthetic complexes, e.g. PSI, PSII and ribulose 1,5-bis-phosphate carboxylase/oxygenase (RuBisCO), were previously suggested to be exclusively transcribed by PEP (Hajdukiewicz, et al., 1997). However, TSSs for numerous of these photosynthesis-related genes in white tissue

were discovered by dRNA-seq. For example, *rbcL* (coding for RuBisCO large subunit) was found to be transcribed by a typical NEP promoter with an YRTa motif in white plastids (Figure 14). T*rbcL*-214 was found enriched after the TEX treatment in white dRNA-seq libraries (Figure 14A). Additionally, 5'-RACE analysis verified the occurrence of this primary transcript in an independent RNA preparation from white plastids (Figure 14B). Furthermore, an additional TSS shortly downstream of T*rbcL*-214 and not detected by dRNA-seq was discovered by 5'-RACE analysis. (Figure 14C). TSSs for the majority of the genes encoding photosystem I and II subunits, among them *psaA*, *psaI, psaJ*, *psbA, psbB*, *psbD*, and *psbM,* were also detected in white tissue. Moreover, the *psaJ* (coding for PSI subunit IX) NEP promoter was even among the most active promoters in white plastids (Appendix A-2).

3.2.6.6 TSSs of *trn* genes

The majority of *trn* genes with TSSs mapped by dRNA-seq in both plastid types were transcribed from upstream promoters (Appendix A and D). In addtiion, a more than two fold enrichment of potential primary transcripts starting exactly at position +1 (relative to gene start) was observed for 6 and 10 *trn* genes in green and white libraries, respectively. Five of them were identical in both plastid types (Appendix C and D). For example, two TSSs were detected for *trnE* in green plastids by dRNA-seq: T*trnE*-46 and T*trnE*+1 (Figure 15A). The upstream TSS (T*trnE*-46; black arrow) is in agreement with the one mapped in *Arabidopsis*. Moreover, the upstream promoter of *trnE* (Figure 15C; red letters) was shown to be dependent on SIG2, a sigma factor of PEP (Hanaoka, et al., 2003). On the other hand, 5'-RACE analysis (Figure 15B) could not discriminate between T*trnE*+1 (Figure 15A; black arrow with an asterisk) being a primary or processed transcript in independent RNA preparation from green plastids (Figure 15 C). In general, the +1 of a *trn* gene is the mature 5' end generated by RNaseP activity, *i.e.* via processing (Thomas, et al., 2000; Wang, et al., 1988). Therefore, the nature of T*trn*+1 mapped by dRNA-seq is rather unclear. These TSSs could result from promoters located directly upstream of the *trn* genes. Another possibility is that the detected T*trn*+1 are generated via transcription from internal promoters as already proposed for several chloroplast tRNA genes in spinach (Cheng, et al., 1997; Gruissem, et al., 1986) and mustard (Liere and Link, 1994; Neuhaus and Link, 1990; Nickelsen and Link, 1990). They might be as well "false positive" TSS candidates

due to the particular structure of tRNAs in which the amino acid acceptor stem might protect some mature 5' ends of tRNAs from degradation by TEX.

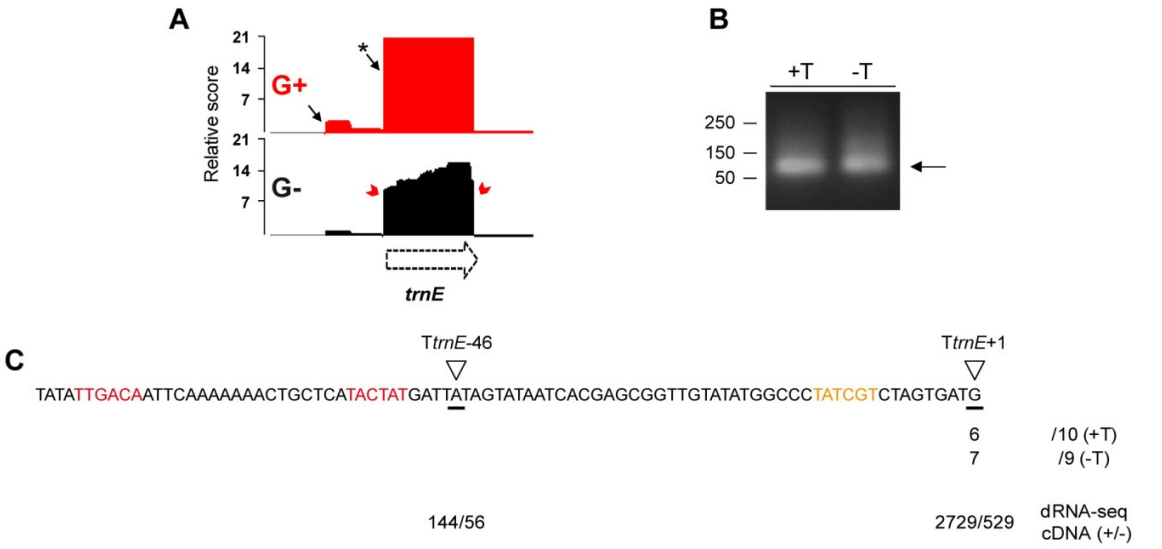


Figure 15: TSSs of *trnE* detected by dRNA-seq. (A) cDNA reads of green (G+/-) dRNA-seq libraries mapped onto *trnE*. T*trnE*-46 (black arrow) and T*trnE*+1(black arrow with an asterisk) were detected. **(B)** 5'RACE TAP treated (+T) and untreated (-T) reactions separated on a 1.5% TAE gel. **(C)** Upstream promoter region of *trnE* used in green plastids. The TSSs identified by dRNA-seq are marked by triangles. The sequence motifs upstream T*trnE*-46 and T*trnE*+1 are shown in red and orange, respectively. The number of clones supporting the 5' end mapping to +1 (relative to gene start) in T+ and T- reactions, as well as the total number of sequenced clones (number after the slash) are given. The numbers of cDNAs in (+/-) dRNA-seq are also listed.

3.2.6.7 TSSs classified as both gTSSs and iTSSs

Numerous TSSs within annotated genes (iTSSs) were detected by dRNA-seq in RNA samples of both green and white plastids (Appendix A). 20 iTSSs were found in green plastids, 7 of which qualified also as TSSs of downstream genes in operons, suggesting a potential role of these TSSs in uncoupling transcription units. Indeed, among them is the previously identified T*psbC*-194 (Sexton, et al., 1990). Ten out of the detected 52 iTSSs in white plastids could also be considered as potential TSSs of downstream genes. For example, an iTSS was mapped within the *rpoB* gene by dRNA-seq in white leaves (Figure 16A). This TSS is located 599 nt upstream of the *rpoC1* start codon and might therefore serve as a gTSS for the transcription of *rpoC1*. 5'-RACE analysis detected T*rpoC1*-599 in samples from both green and white plastids (Figure 16B). Moreover, an YRTa motif was detected upstream of T*rpoC1*-599 (Figure 16C). Taken

together, these observations suggest that T*rpoC1*-599 is an operon- and *rpoB*-internal TSS allowing for separate transcription of *rpoC1* in both green and white leaves.

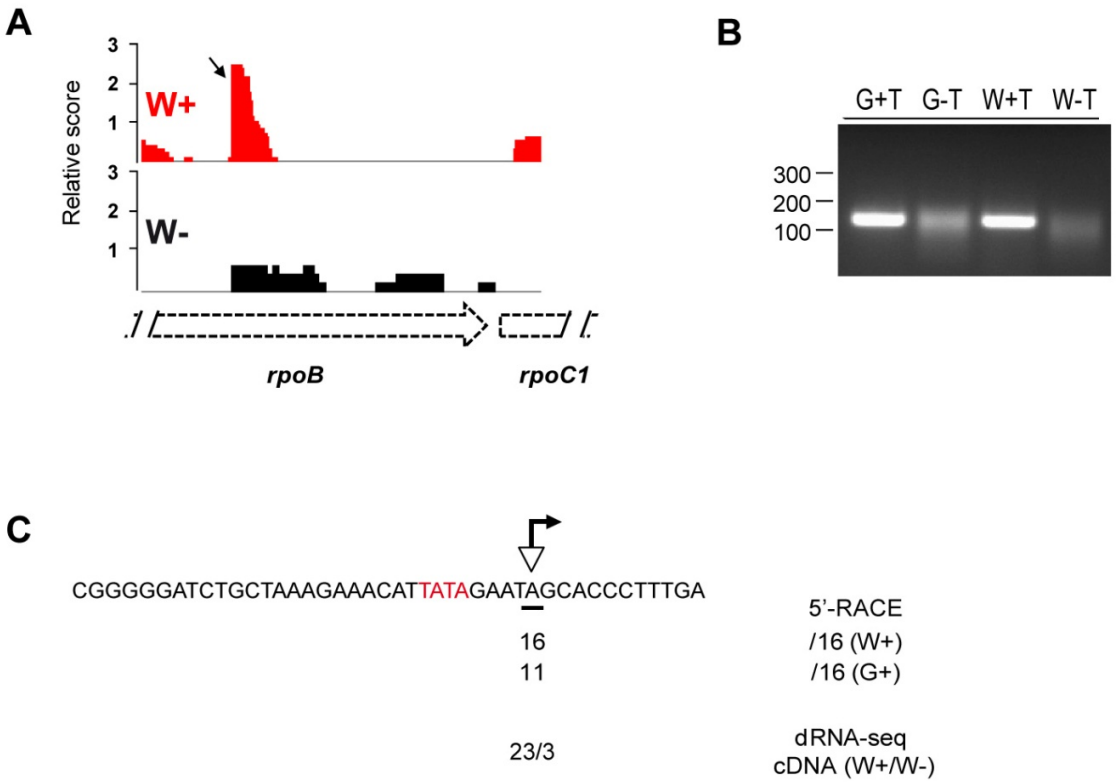


Figure 16: T*rpoC1*-599 - both an iTSS and a gTSS. (A) cDNA reads of white (W+/-) dRNA-seq libraries mapped onto *rpoB-C1*. T*rpoC*-559 (black arrow) was detected within the *rpoB* ORF. **(B)** 5'-RACE revealed the presence of T*rpoC*-559 in both green (G) and white (W) plastids. 5'RACE TAP treated (+T) and untreated (-T) reactions separated on a 1.5% TAE gel. The products of expected size are marked by an arrow. **(C)** Upstream promoter region of T*rpoC*-559. The TSS identified by dRNA-seq and 5'-RACE are marked by a triangle and arrows, respectively. The sequence motif upstream T*rpoC*-559 is shown in red. The number of clones supporting T*rpoC*-559 in W+ and G+ reactions, as well as the total number of sequenced clones (number after the slash) is given. The number of cDNAs in (W+/W-) dRNA-seq is also listed.

3.2.7 Potential regulatory ncRNAs in plastids

One of the main questions that were addressed in this study was the occurrence of ncRNAs in plastids. Indeed, numerous aTSSs and oTSSs were mapped by dRNA-seq in barley *albostrians* plastids (Figure 9, Appendix A), indicating the synthesis of many ncRNAs. Fifty-five aTSSs were detected in both green and white plastids leading to the synthesis of antisense transcripts to 40 and 45 genes, respectively. Seventeen and 30 of the TSSs in green and white plastids, respectively, were mapped to intergenic regions, and thus assigned as oTSSs.

Based on the set of aTSSs and oTSSs mapped in green libraries, 60 ncRNA candidates (denoted Hv_nc) generated by transcription of free-standing genes were predicted in mature barley chloroplasts (Appendix E). aTSSs that were also annotated as gTSSs were not considered as TSSs of ncRNA candidates. Several aTSSs were mapped in intergenic regions, and therefore assigned to both the upstream and downstream encoded RNA. Hv_nc3 (T*psbK*-783), Hv_nc39 (as_*rps18*; T*trnP*-1937) and Hv_nc46 (as_*petD*; T*psbN*-3371) might be transcribed via a NEP promoter in chloropalsts, since their TSSs were found identidal in both G and W dRNA-seq libraries (Appendix C and E; see Chapter 3.2.6.4. for more details on identical TSSs in G and W libraries). Moreover, 5'-RACE analysis detected two potential ncRNAs, Hv_nc56 (as_*rps15;* T*trnN*-1479) and Hv_nc58 (as_*ndhF*; T*rpl32*-1224), the TSSs of which were solely visible in white dRNA-seq libraries, to be also present in green plastids with the same initiating nucleotide, thus revealing them as potential NEP-dependent ncRNA candidates in chloroplasts (Appendix B). Consequently, it is possible that more of the aTSSs or oTSSs exclusively found in white libraries are actually TSSs of ncRNAs that are also expressed in green plastids. In addition, some of the more distant gTSSs could be used for ncRNA synthesis rather than expression of annotated genes. It cannot be ruled out that the TSSs of some of the predicted ncRNAs could be driving transcription of annotated genes located further downstream. Therefore, seven ncRNA candidates were selected and their 3' ends were mapped precisely by 3'-RACE. Indeed, the 3'-RACE analysis verified the synthesis of ncRNAs (of size between 50 and 400 nt) from the corresponding TSSs (Appendix E).

Forty-eight (80%) of the ncRNA candidates in mature barley chloroplasts presented in this study could be classified as potential *cis*-encoded asRNAs (Appendix E). asRNAs were found complementary to both photosynthesis and genetic system genes. Moreover, the predicted *cis*-encoded asRNAs in this study were targeting either the coding region or different non-coding parts of plastid genes (Table 5). The TSSs of 14 ncRNA candidates were mapped within annotated ORFs on the opposite strand. The TSSs of 8 and 10 of the ncRNAs were complementary to 5' and 3' UTRs, respectively. In addition, the TSSs of 3 and 6 ncRNA were complementary to 5' and 3' regions, respectively, of *trn* or *rrn* precursors. Moreover, the TSSs of 12 ncRNAs were mapped within introns on the opposite strand (Table 5). An example of an *cis*-encoded ncRNA is presented in Figure 17. dRNA-seq detected the primary end of an asRNA (Hv_nc40) against petB exon1/intron region (Figure 17A). *as_petB* is complementary to the

intron-binding sites 1 and 2 of the *petB* introns which is a group II intron (Michel and Ferat, 1995).The 5' and 3' ends of the ncRNAs, and some other selected candidates, were additionally determined by 5' and 3'-RACE analysis (Figure 17B and C; Appendix B and E).

Table 5: Targets of *cis*-encoded asRNA candidates predicted in this study

Target regions	Target genes[1]
ORF	*psbA*, *psbD*, *rpoC1*, *rpoC2*, *atpA*, *psaA*x2, *rbcL*x2, *petB*, *ndhB*, *ndhF*x2, *ndhD*
5' UTR	*rps14*, *ndhD*, *psbL*, *psbC*, *rps16*, *atpI*, *psaA, rps15*
3' UTR	*petG*, *petD*, *ycf3*, *rbcL*, *psaJ*, *rps18*, *psaC*, *psaB*, *psbF, psbM*
5' region	*trnE*, *trnY, trnT* precursors
3' region	*trnF*x2, *trnL*, *rrn5*, *trnR*, *trnE* precursors
intron	*trnG*x2*, atpF, ycf3, trnV-UAC, petB*x3*, petD*x2*, rpl2, trnA*

[1]In several cases, more than one asRNA (marked by x2/x3) were mapped per annotation and a singular asRNA was assigned to two targets (not shown).

Twelve of the ncRNA candidates in this study were found in intergenic regions, and thus could have a potential function as *trans*-encoded asRNAs, i.e. act on one or more unlinked *trans*-encoded targets through short rigions of complementarity. Bacterial *trans*-encoded asRNAs often base pair with the ribosomal binding site (RBS) region of the target mRNA, and thus block its translation (Vogel and Wagner, 2007). The maximal region covered by ribosomes was experimentally determined to be -39 to +19 relative to the start codon (Huttenhofer and Noller, 1994). Here, intaRNA (Busch, et al., 2008; Smith, et al., 2010), a software for computational prediction of the interaction of two RNA molecules, was used to screen with default parameters mRNA regions from -50 to +25 relative to the start codon for potential targets of the identified ncRNAs from intergenic regions. In total, 99 candidates with predicted energy of interaction <-8.5 kcal/mol were identified (data not shown). The top 10 highest scoring candidates (energy <-14.5 kcal/mol) are given in Appendix F. It will be worth experimentally testing if some of them are real targets of the corresponding ncRNAs.

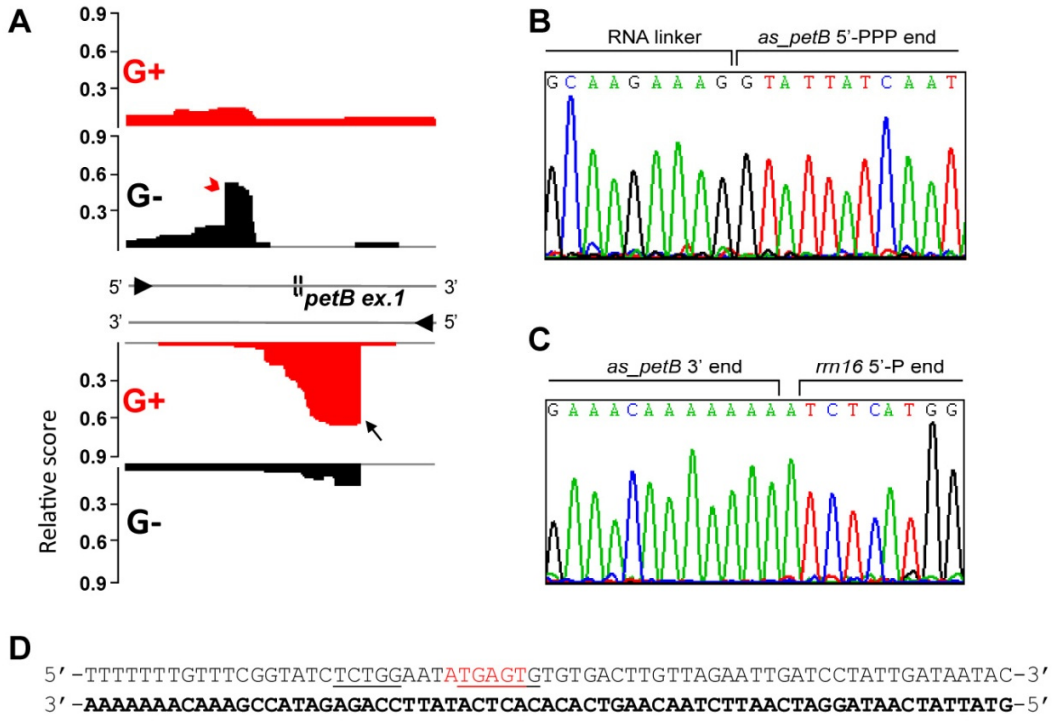


Figure 17: A non-coding RNA candidate in mature plastids. (A) cDNA reads of green (G+/-) libraries mapped onto *petB* exon 1. The detected 5' processed end of *petB* is indicated by a red arrow head and the TSS of the *petB* antisense transcript, as_*petB* (Hv_nc40) is marked by a black arrow. The direction of transcription is indicated by black arrow heads on the DNA strands. **(B)** A Chromatogram displaying sequence at the ligation site of a cloned 5'-RACE product from the T+ reaction. The 3' end of the linker and the 5'-PPP end of the *as_petB* are shown. **(C)** A Chromatogram displaying sequence at the ligation site of a cloned 3'-RACE product from the ligase + reaction (see Chapter 2.2.12.2). The 3' end of the *as_petB* and the 5'-P end of the rrn16 which was used as a linker are shown. **(D)** *as_petB* sequence. *as_petB* is marked in bold. *petB* exon 1 is colored in red. Intron-binding site 1 and 2 of the *petB* are underlined.

3.3 The processed transcriptome of mature barley chloroplasts

3.3.1 Annotation of processing sites based on dRNA-seq

In total, 31 5'-P ends (processing sites; PSs), susceptible to degradation by TEX were mapped in green *albostrians* plastids (Appendix G). PS*psbD*-135, PS*psbC*-9, PS*trnT*-2006 and PS*psbB*-63 were undetected in white libraries, most likely due to the low abundance of these processed transcripts in white plastids. This part of the study will focus on the PSs mapped in green *albostrians* plastids and will refer to them as PSs of mature barley chloroplasts. The criteria for annotation of PSs are summarized in the Materials and Methods section: Annotation of TSSs and PSs based on dRNA-seq. Mature 5' ends of tRNAs and rRNAs were not included in the list

(Appendix G). Two of the PSs, PS*rrn16*-28 and PS*rrn23*-73 (PS are named after the downstream located gene and the number of nucleotides between the 5'-P and the start codon of the ORF) were assigned as 5'-P of 16S and 23S rRNA precursors, respectively. Six PSs were mapped in closer proximity to upstream genes and most likely do not stand for 5' ends of downstream RNAs (marked in the Comments section of Appendix G). One PSs, PS*atpA*-1411, was mapped with *atpF* exon 1, and thus its attribution is unclear. The remaining 22 PSs were mapped shortly upstream of annotated protein-coding genes, and thus with high probability represent processed 5' ends of the corresponding mRNAs. Indeed, the cDNA reads of the majority of them reached the downstream located ORFs. Moreover, eleven of the 22 PSs were in agreement with previously published data on chloroplast processed 5' mRNAs ends. Additionally, nine of them were verified as 5'-P ends in independent RNA preparations from barley chloroplasts by 5'-RACE analysis using primers mapping within the corresponding downstream ORFs (Appendix H).

3.3.2 Small RNAs associated with processed 5' mRNA ends revealed by dRNA-seq

3.3.2.1 Small RNAs matching known or predicted binding sites of PPR proteins

Maize PPR10 was shown to participate in the formation of 5' and 3' processed mRNA termini mapping to the intergenic regions of *atpH-atpI* and *psaJ-rpl33* by serving as a barrier to 5' to 3' and 3' to 5' exonuclease activity, respectively (Pfalz, et al., 2009; Prikryl, et al., 2011). Moreover, small RNAs (sRNAs) corresponding to the PPR10 binding sites were detected in the sRNA transcriptome of rice and maize (Johnson, et al., 2007; Morin, et al., 2008; Pfalz, et al., 2009). It was proposed that these sRNAs stand for *in vivo* footprints of PPR10, i.e. the minimal PPR binding sites protected from complete ribonuclease attack. Indeed, the 5' and 3' termini of the sRNAs harboring the PPR10 binding site correspond with the positions at which 5'- and 3'-exonucleases are stalled by recombinant PPR10 *in vitro* (Prikryl, et al., 2011).

A main goal of this study was to investigate the occurrence of sRNAs representing footprints of bound PPR proteins in the transcriptome of mature barley chloroplasts. These sRNAs should correspond with processed 5' mRNA termini mapped in this study. Indeed, a closer look at the 5'-P ends of *atpH* (PS*atpH*-49) and *rpl33* (PS*rpl33*-161) mapped by dRNA-seq revealed their association with ~25 nt long RNAs which matched perfectly the described PPR10 binding sites in maize (Figure 18; Pfalz, et al., 2009; Prikryl, et al., 2011). Moreover, sRNAs corresponding to

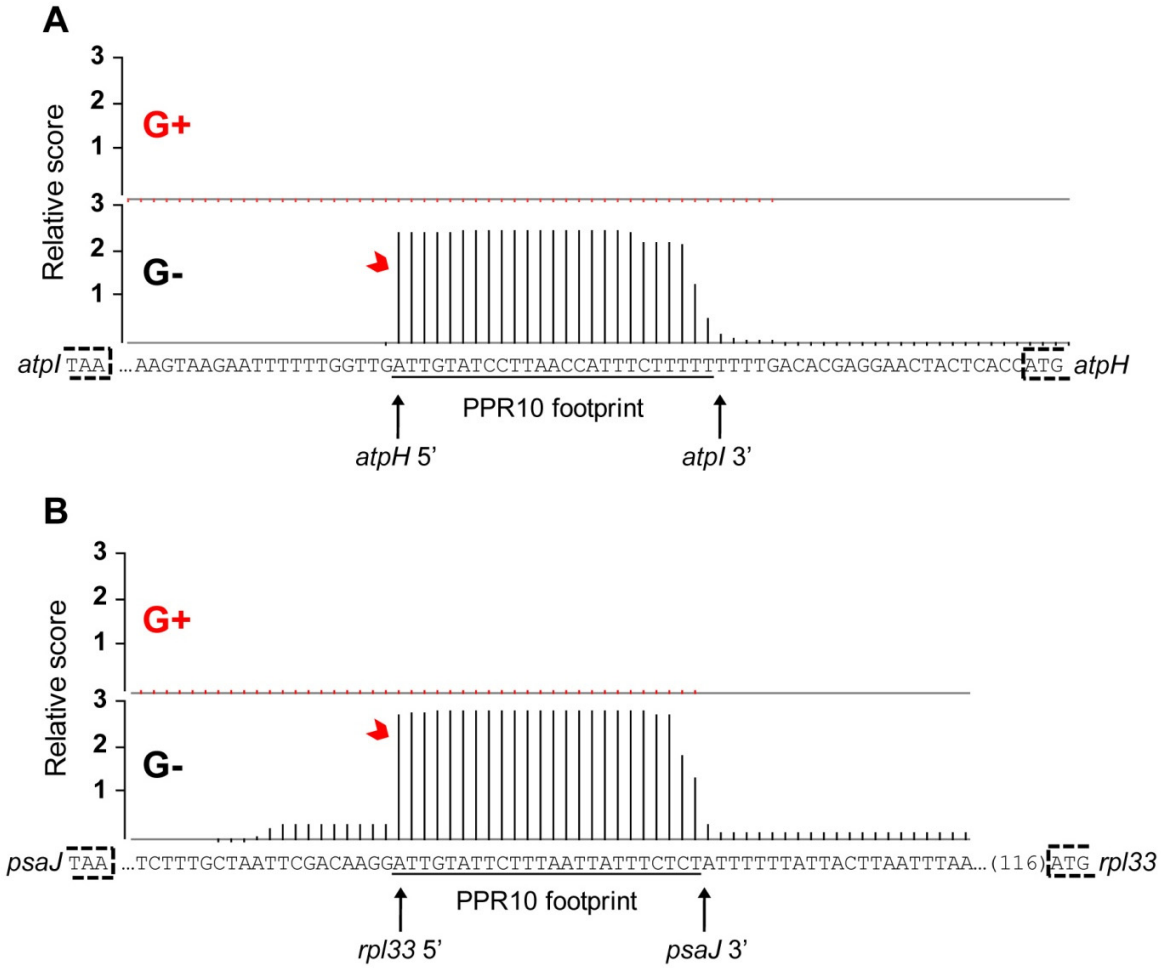


Figure 18: Small RNAs representing PPR10 footprints revealed by dRNA-seq. (A) cDNA reads of G+/G- libraries mapped onto the *atpI-atpH* intergenic region. **(B)** cDNA reads from G+/G- libraries mapped onto the *psaJ-rpl33* intergenic region. Small RNAs, matching the previously defined PPR10 binding sites, was found associated with *atpH* and *rpl33* mature 5' end (PS*atpH*-49 and PS*rpl33*-161; red arrow heads). The PPR10 biding sites and the 5' and 3' termini stabilized by this protein are marked.

the binding sites of two other PPR proteins were detected (Table 6). PG3 stabilizes *petL* 5'terminus (Yamazaki, et al., 2004), and HCF152 is required for the accumulation of RNAs with processed 5' and 3' ends mapping to the *psbH-petB* intergenic regions (Meierhoff, et al., 2003). Here, sRNAs corresponding to the mapped binding sites of PGR3 (Cai, et al., 2011) and HCF152 (Zhelyazkova, et al., 2011) were found to be associated with PS*petL*-66 and PS*petB*-44, respectively.

sRNAs were also detected at several processed ends shown to be sites of action of other genetically characterized PPR proteins, but for which no exact binding site has been determined

Table 6: sRNAs associated with mature 5' mRNA ends in barley chloroplasts matching known or predicted binding sites of PPR proteins

Mature 5' mRNA end	Region	Small RNA sequence[1]	Species[2]	Protein[3]	RNA termini[4]
PS*atpH*-49	*atpI-atpH*	ATTGTATCCTTAACCATTTCTTTT	Hv, Zm, At, Os	PPR10	*atpI* 3' *atpH* 5'
PS*rbcL*-59	*rbcL* 5'	CATCGAGTAGACCCTGTTATTGTGAGAATT	Hv, Os, At	MRL1*	***rbcL* 5'**
PS*petL*-66	*petL* 5'	CTTAGGTAAATGCTTTACCAACATATGTAGT	Hv, At, Os	PGR3	***petL* 5'**
PS*rpl33*-161	*psaJ-rpl33*	ATTGTATTCTTTAATTATTTCTCT	Hv, Os	PPR10	*psaJ* 3' *rpl33* 5'
PS*rps12*-52	*clpP-rps12 5'*	ATCAGGTTAAGATGGATCTAAACCAATCCATTTTT	Hv, Zm, At, Os	PPR38*	***clpP3'* *rps12* 5'**
PS*psbH*-37	*psbT-psbH*	AGTATACAAAGTCAACACCAATGATT	Hv, Zm, At, Os	HCF107*	*psbH* 5'
PS*petB*-44	*psbH-petB*	GGTAGTTCGACCGCGGAATT	Hv, Zm, At, Os, Nt	HCF152	*psbH*3' *petB* 5'
PS*petD*-148	*petB-petD*	ATATCGGGTAGGTTGTGGTATTTCATTGCT	Hv, Zm, Os	CRP1*	*petB* 3' *petD* 5'
PS*ndhB*-17	*rps7-ndhB*	ATGCAGTTACTAATTCATGATCTGGCATGT	Hv, Zm, At, Os, Nt	CRR2*	***rps7* 3'** *ndhB* 5'

[1]T is shown instead of U in the sequences of the sRNAs. Each sRNA represents a population of molecules with ends mapping within several nucleotides of the given sequence.

[2]The species in which the sRNAs have been reported are listed as follows: Hv- *Hordeum* vulgare (data from this study); Zm- *Zea mays* (data from http://sundarlab.ucdavis.edu/smrnas/, Johnson, et al., 2007); Os- *Oryza sativa* and At- *Arabidopsis thaliana* (data from Ruwe and Schmitz-Linneweber, 2011); Nt- *Nicotiana tabacum* (data from Lung, et al., 2006).

[3]Proteins that are predicted (marked with asterisk) or have been shown to bind to the corresponding sequences are given. References are provided in the main text.

(Table 6; marked by asterisk). Thus, these sRNAs might represent the *in vivo* footprints of the corresponding proteins. For example, the abundant sRNA associated with PS*petD*-148 corresponded to the sequence shared by the 5' of *petD* and 3' of *petB* processed mRNAs, shown to be CRP1-dependent transcripts (Barkan, et al., 1994; Fisk, et al., 1999). Equivalently, PS*rps12*-52 mapped the 5' of a sRNA which boundaries matched the processed 5'- and 3'-termini in the *clpP-rps12* intergenic region in maize proposed to be stabilized by PPR38 (Hattori, et al., 2007; Hattori and Sugita, 2009; Zhelyazkova, et al., 2011). Another sRNA, associated with PS*rbcL*-59, corresponded to the site of action of the *Arabidopsis* PPR protein MRL1, required for the stabilization of *rbcL* 5' (Johnson, et al., 2010). PS*psbH*-37 and PS*ndhB*-17, HCF102- and CRR2- dependent 5' ends in *Arabidopsis* (Sane, et al., 2005; Hashimoto, et al., 2003), respectively, were also associated with sRNAs. In addition, 3'-RACE analysis revealed that the 3' end of the sRNA representing the putative CRR2 binding site matched also the 3' end of the

upstream located *rps7* (Appendix H). All the sRNAs described above have orthologs in at least one other species (Table 6; Johnson, et al., 2007; Lung, et al., 2006; Ruwe and Schmitz-Linneweber, 2011).

3.3.2.2 Small RNAs are associated with the majority of processed 5' mRNA ends mapped by dRNA-seq

In total, 21 of the 22 processed 5' mRNA ends mapped in this study were found to be associated with sRNAs. Nine of these sRNA corresponded to predicted or known binding sites of PPR proteins (Table 6). Moreover, it was demonstrated that the maize orthologs of the sRNAs mapping in the *atpH-atpI* and *petB-petD* intergenic region were indeed PPR10 and CRP1-dependent features, respectively (Zhelyazkova, et al., 2011). In parallel, HCF152 was shown to be required for the accumulation of the *Arabidopsis* ortholog of the *psbH-petB* sRNA (Ruwe and Schmitz-Linneweber, 2011). Taken together, these results strongly support the view that the sRNAs matching the genetically defined targets of PPR proteins are indeed *in vivo* footprints of the bound proteins. Therefore, the 12 sRNAs, associated with 5'-P mRNA ends mapped by dRNA-seq for which no stabilizing protein has been reported, might also represent binding sites of PPR or PPR-like proteins (Table 7). In general, highly structured RNA stretches seem unlikely to serve as binding sites for PPR proteins (Prikryl, et al., 2011; Williams-Carrier, et al., 2008). Moreover, these RNAs might be nuclease-resistant due to their stable secondary structure, rather than due to protection by a bound protein. Only the sRNAs associated with PS*ndhA*-67 and PS*ndhB*-173 had the tendency to form a fairly stable stem loop structures (dG=-10.7 kcal/mol and dG=-11.8 kcal/mol, respectively; Zhelyazkova, et al., 2011). However, the processed 5' end of *ndhA* and an ortholog of the associated sRNA are also found in *Arabidopsis* (Ruwe and Schmitz-Linneweber, 2011) and are in agreement with a proposed binding site for the PPR protein PGR3 (Cai, et al., 2011). Moreover, the boundaries of the PS*ndhB*-173 sRNA correlate with the 5' end of *ndhB* mapped here by 5'-RACE (Appendix H) and the *rps7* 3' end reported in *Arabidopsis* (Hashimoto, et al., 2003). This hints for these sRNAs also reflecting binding sites for stabilizing proteins. Thus, both RNA-structure and protein-protection mechanisms are equally possible to participate in the stabilization of the 5' end of *ndhA* and *ndhB* processed transcript (and the 3' of the corresponding *rps7* transcript). The other 10 sRNAs lack stable secondary structure, and are therefore good candidates for RNAs protected by a bound protein.

Moreover, many of them are accompanied by orthologous sRNAs in at least one other species (Table 7). In addition, all sRNAs except two, the one associated with PS*rps2*-97 and PS*ycf3*-62 (conserved only among monocots), are highly conserved between monocots and dicots (Zhelyazkova, et al., 2011). Thus, at least 19 out of the 22 processed 5' mRNA ends mapped by dRNA-seq in this study are associated with sRNAs which are excellent candidates for *in vivo* footprints of PPR or PPR-like proteins. That is, these 5' termini are with high probability generated *via* protection from 5' to 3' degradation by a bound protein.

Table 7: sRNAs associated with mature 5' mRNA ends for which no stabilizing proteins have been identified

Mature 5' mRNA end	Region	Small RNA sequence[1]	Species[2]	RNA termini[3]
PS*rps16*-79	*rps16*' 5'	AAACCAATGACTATTCATGATTCCATCCAT	Hv, At, Os, Nt	***rps 16* 5'**(Hv)
PS*psbC*-45	*psbC* 5'	ATCAGCCTCATGAAAATCTTATATA	Hv, At, Os, Zm	*psbC* 5' (Hv, At, Zm)
PS*rpoB*-126	*rpoB* 5'	TAGAATTTCATGTGATTCAGTAAACAGAATATA†	Hv, Os	*rpoB* 5' (Hv)
PS*rps2*-97	*rps2* 5'	ATTTATTTCAAGCTATTTCGGATCTT	Hv, Os	*rps2* 5' (Hv)
PS*rps14*-59	*rps14* 5'	ATTTATTTTTCCATCTAGGATTAGAACCGTATACT	Hv	*rps14* 5' (Hv)
PS*ycf3*-62	*rps4-ycf3*	TTTGTTTTTATGTTATTTTGTGAAG	Hv, At, Os,	***rps4* 3'** (Hv); ***ycf3* 5'** (Hv)
PS*ndhK*-57	*ndhK* 5'	TTTCGTGCTTATCTTAGTTGTCGGTTTAGT	Hv	***ndhK* 5'** (Hv)
PS*psbB*-63	*psbB* 5'	TTTTTCAATGCGATAAAATAAAGCGACATCGTGT	Hv, At, Os	***psbB* 5'** (Zm,Hv,At)
PS*ndhB*-173	*rps7-ndhB*	GAAATCATGATCAACTAAGCCCTCTCGAGGGCTT*	Hv	*rps7* 3' (At); ***ndhB* 5'** (Hv)[4]
PS*rps12*-683	*rps12* 5'	CAACATAGGTCATCGAAAAGATCTCGGACAACTCA	Hv	5' end of 2nd fragment (Hv)
PS*psaC*-188	*ndhE-psaC*	CAAAATTCAAGTCTCTTGGCTCTTTTCACGC	Hv, At	***ndhE* 3'** (Hv); ***psaC* 5'** (Hv)
PS*ndhA*-67	*ndhA* 5'	AAATTGGCTGATATCATGACGATATTAGGTAG*	Hv, At	*ndhA* 5' (Hv, Zm, At)

[1]T is shown instead of U in the sequences of the sRNAs. Each sRNA represents a population of molecules with ends mapping within several nucleotides of the given sequence. sRNA with predicted stable secondary structure are marked with an asterisk. The sRNA indicated by a dagger was found in close proximity to the TSS T*rpoB*-147.

[2]The species in which the sRNAs have been reported are given as in Table 3.

[3]The indicated 5' and 3' RNA termini match 5' and 3' termini of the corresponding sRNA. Species in which the end has been detected are indicated in parentheses. The RNA termini verified here by 5'/3'-RACE or cRT-PCR are marked in bold (Appendix G).

[4]Two sets of 5' and 3' termini are mapped in the *rps7-ndhB* intergenic region. The termini depicted here are the ones spatially closer to the *rps7* ORF.

The sRNA associated with PS*ycf3*-62 is a good example of a potential protein binding site revealed by dRNA-seq (Figure 19A). This small unstructured RNA is among the most abundant ones detected in this study (Appendix G). Moreover, its orthologs are also found in rice and *Arabidopsis* (Ruwe and Schmitz-Linneweber, 2011). The processed 5' end of *ycf3* was

additionally verified here by 5'-RACE (Figure 19B). Furthermore, 3'-RACE analysis revealed that the 3' of PS*ycf3*-62 sRNA corresponded with the 3' ends of the upstream transcript *rps4*, providing evidence for bidirectional RNA stabilization by a protein bound to this site (Figure 19C).

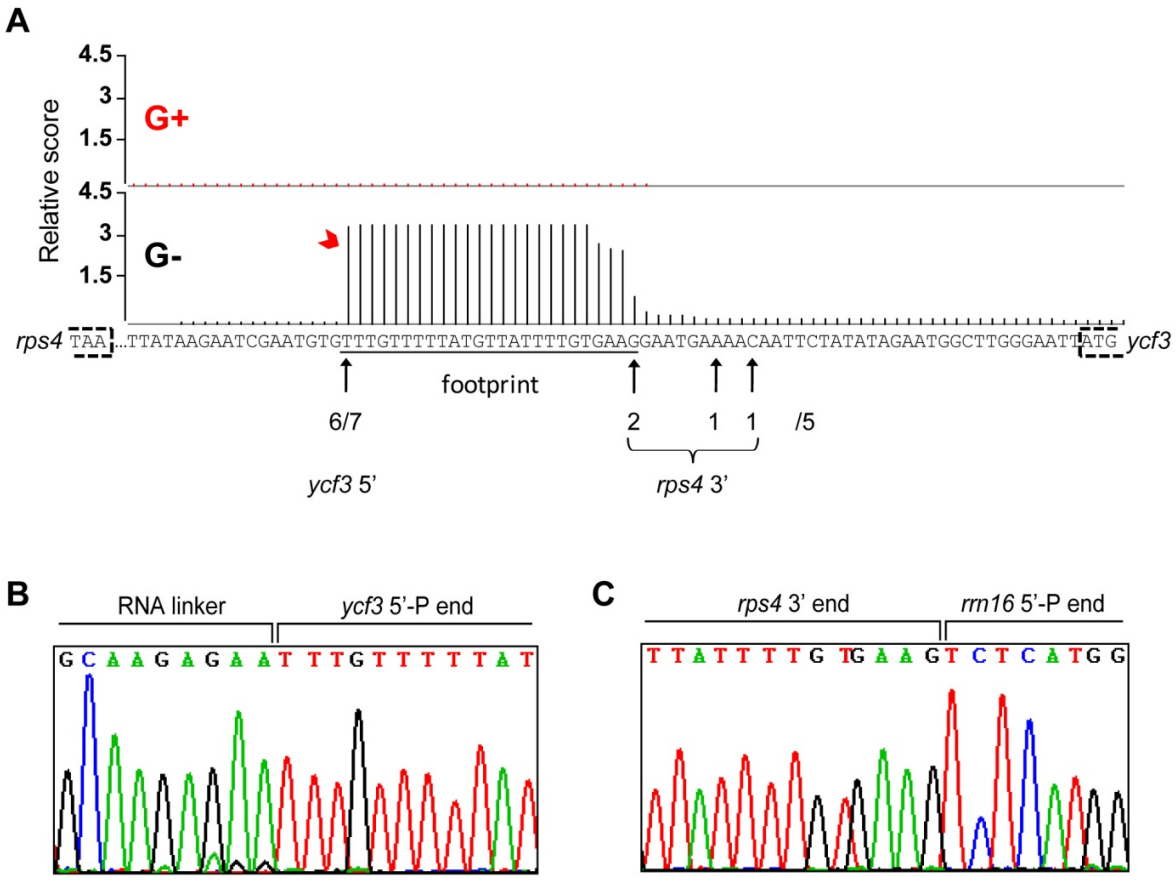


Figure 19: The sRNA associated with PS*ycf3*-62 - a potential footprint of an uncharacterized PPR/PPR-like protein. (A) dRNA-seq cDNA reads of G+/G- libraries mapped onto the *rps4-ycf3* intergenic region. PS*ycf3*-62 (red arrow head) is associated with a sRNA which lacks secondary structure (data not shown) and most likely represents the *in vivo* footprint (underlined) of a bound PPR/PPR-like protein. The boundaries of the sRNA match the *ycf3* 5' and *rps4* 3' ends mapped by 5'- and 3'-RACE, respectively. The number of clones supporting each termini, as well as the total number of sequenced clones (number after the slash) are given. **(B)** A Chromatogram displaying sequence at the ligation site of a cloned 5'-RACE product from the T- reaction. The 3' end of the linker and the 5'-P end of the *ycf3* are shown. **(C)** A Chromatogram displaying sequence at the ligation site of a cloned 3'-RACE product from the ligase+ reaction (see Chapter 2.2.12.2). The 3' end of the *rps4* and the 5'-P end of the *rrn16* which was used as a linker are shown.

Another interesting example is the sRNA associated with PS*ndhK*-57, i.e the 5'-P end of *ndhK,* which is found within the upstream *ndhC* ORF (Figure 20). The processed 5' end of *ndhK*

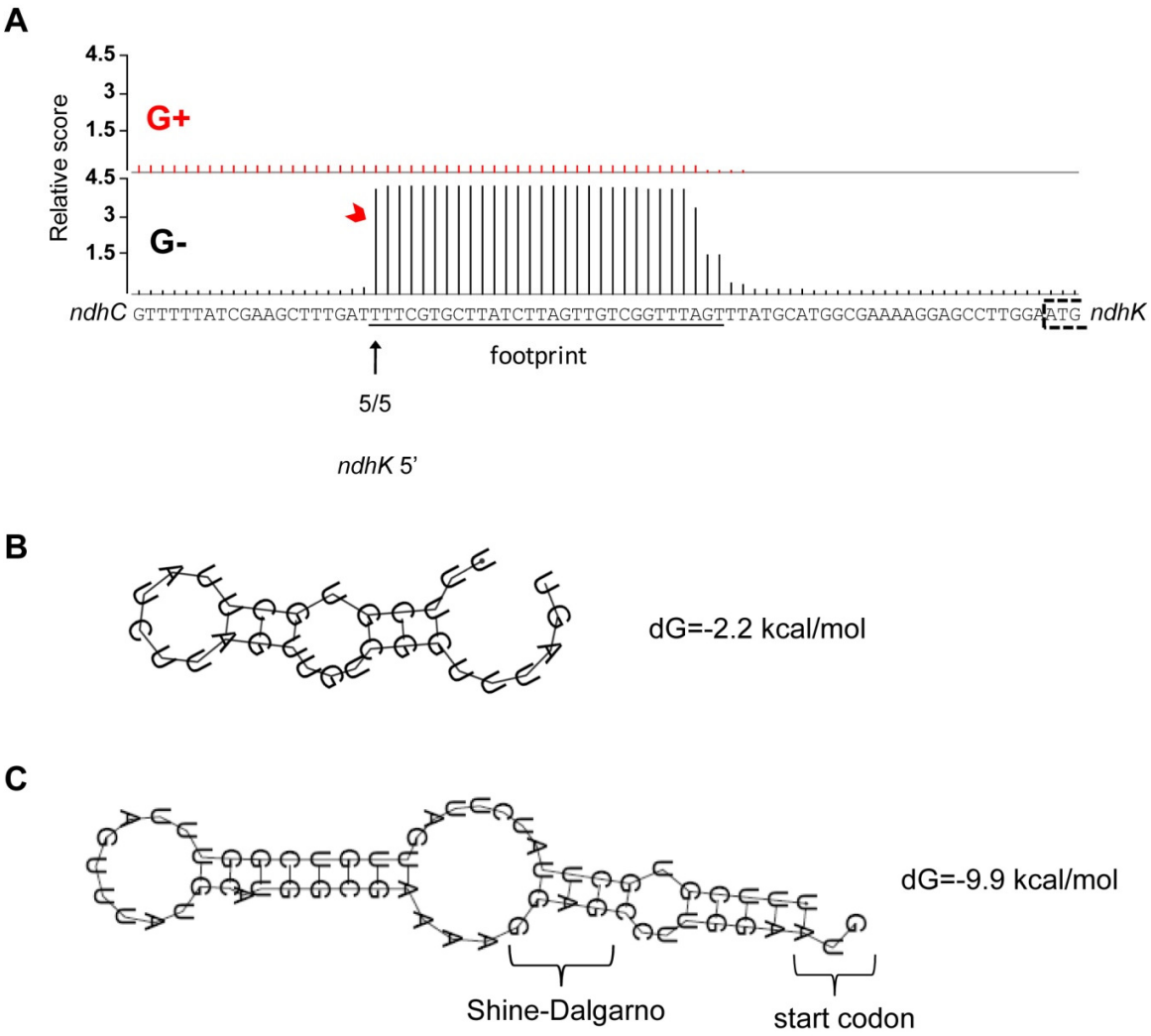


Figure 20 Binding of a putative PPR to *ndhK* 5' UTR could result in translational enhancement. **(A)** dRNA-seq cDNA reads of G+/G- libraries mapped onto the *ndhC-ndhK* region. PS*ndhK*-57 (red arrow head) is associated with a sRNA most likely representing the *in vivo* footprint (underlined) of a yet uncharacterized PPR/PPR-like protein. The 5'-P end of *ndhK* was additionally verified by 5'-RACE. The number of clones supporting the 5' terminus, as well as the total number of sequenced clones (number after the slash) is given. **(B)** Secondary structure prediction of the sRNA associated with PS*ndhK*-57. **(C)** Secondary structure prediction of *ndhK* 5'-UTR region. If not occupied by a protein, the putative PPR binding site can base pair with the Shine-Dalgarno sequence (marked) resulting in the formation of a stable structure. Secondary structure predictions were made using RNAfold Server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

mapped by dRNA- seq was additionally verified here using 5'-RACE analysis (Figure 20A). The sRNA associated with PS*ndhK*-57 is an excellent candidate for a PPR footprint since it was predicted to lack a stable secondary structure (Figure 20B). Interestingly, a structure prediction of the *ndhK* 5' UTR region revealed that the putative PPR binding site could base pair with the

the Shine-Dalgarno sequence resulting in the formation of a stable structure (dG=-9.9 kcal/mol; Figure 20C). Therefore, it can be speculated that similar to the PPR10 mode of action at the *atpH* 5' UTR (Prikryl, et al., 2011), binding of a PPR protein to this region would enhance translation by hindering the formation of the secondary structure. The same scenario is also possible upon binding of a PPR protein to the *psbC* 5'UTR (Zhelyazkova, et al., 2011).

PS*psbD*-135 is the only processed 5' mRNA end mapped in barley chloroplasts that is not associated with a sRNA. The 5' end of the processed *psbD* transcript is most likely protected from a ribonuclease attack by a stable stem-loop structure (dG=-12.6 kcal/mol) predicted to form within the first 50 nt of this mRNA. PS*rrn16*-28 and PS*rrn23*-73, the 5'-P ends of 16S and 23S rRNA precursors mapped by dRNA-seq, respectively, were also not associated with sRNAs.

3.3.3 Processed 3' mRNA ends revealed by dRNA-seq

3.3.3.1 Processed mRNA 3' ends defined by stable stem-loop structures

Processed transcripts associated with stable structures are protected from TEX digestion (Sharma, et al., 2010). The same was observed here in the dRNA-seq profile of *psbA* from green chloroplasts (Figure 21). After the TEX treatment, a distinctive stepwise accumulation of cDNAs in proximity to the 3' end of the *psbA* ORF was observed in both the (-) and (+) libraries, but more pronounced in the latter (Figure 21A; red arrowheads). Moreover, this cDNA accumulation was characterized by a sharp 3' end, rather than a sharp 5' end as in the case of TSSs (black arrow). A comparison with data available in the literature revealed that the majority of the 3' ends of these cDNA reads matched precisely the last base-pair of the stem-loop structure formed at the 3' end of the *psbA* mRNA (Figure 21B; Memon, et al., 1996). In general, stem-loop structures can participate in mRNA 3' end formation by stalling 3' to 5' exonuclease activity (Stern, et al., 2010). In total, comparable TEX-resistant cDNA accumulations were mapped near the 3' ends of 13 genes. These may mark the 3' ends of the corresponding mRNAs. Indeed, two of them were experimentally verified as 3' ends and another three correspond to previously mapped mRNA 3' termini (Appendix I). Furthermore, a high probability of stem-loop structure formation was predicted in all cases. Thus, these 13 potential 3' mRNA ends in barley chloroplasts are proposed to be generated *via* RNA structure-mediated blockage of 3' nucleases.

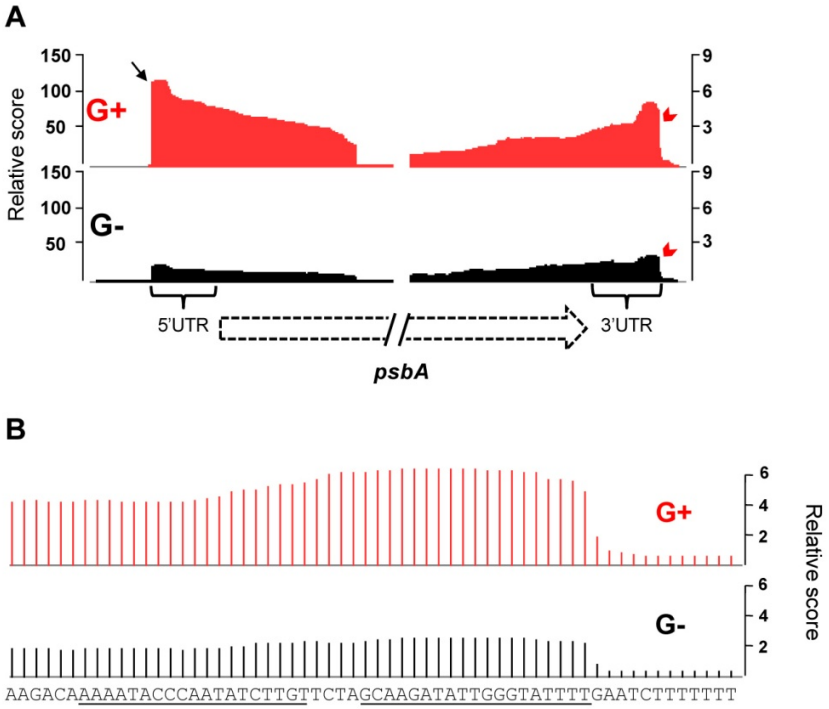


Figure 21 A stable stem-loop structure defines *psbA* mRNA 3' end. **(A)** dRNA-seq profile of *psbA* in green plastids. The TSS of *psbA* (black arrow), is found significantly enriched after the TEX treatment. TEX resistant cDNAs (red arrowhead), corresponding to 3'terminal hairpin RNAs, reveal the 3' end of *psbA* mRNA. The 5' and 3' UTR regions are indicated. **(B)** A close-up view of the cDNA reads of green (G+/-) libraries mapped to *psbA*. A distinctive stepwise accumulation of cDNAs in proximity to the 3' end of the *psbA* ORF was observed to be more pronounced in G+. The most predominant 3' end of these cDNAs matches precisely with the last base-pair of the previously described stem-loop structure (Memon et al., 1996; complementary region is underlined) formed at the 3' end of *psbA* mRNA.

3.3.3.2 Processed mRNA 3' ends defined by protein-mediated protection

An alternative means for mRNA 3' end formation, *i.e.* blocking 3' to 5' exonucleases by a bound protein, has been demonstrated for the PPR10-dependent *atpI* and *psaJ* 3' ends (Pfalz, et al., 2009; Prikryl, et al., 2011). The protein-protection mechanism results in the placement of these 3' termini at the 3' ends of the sRNAs harboring the PPR10 binding sites (discussed in Chapter 3.2.1.). Analogously, the 3' ends of *clpP*, *psbH*, *petB* and *rps7* (long) should result from the same mechanism, since their 3' termini match the 3' ends of the sRNAs which represent with high probability the binding sites of the corresponding genetically characterized PPR proteins (Table 6).In addition, *ndhE* and *rps4* 3' processed termini might also arise *via* protein-protection. Even though no stabilizing proteins have been previously described for these transcripts, a 3'-RACE analysis in this study mapped *ndhE* and *rps4* 3' termini in accordance with the 3' ends of

sRNAs which are excellent candidates of PPR/PPR-like binding sites (Table 7). Due to the high probability for a stable structure formation of the sRNA associated with one of the *rps7* 3' ends (short; see Chapter 3.3.2.2.), it is unclear if this mRNA terminus is protected from 3'-nucleases by a stable structure or a bound protein.

As already mentioned, six 5'-P ends mapped by dRNA-seq were located in closer proximity to upstream genes and were not considered to stand for 5' ends of downstream RNAs. Moreover, they were found associated with small unstructured RNAs (Table 8). These sRNAs could represent footprints of uncharacterized PPR/PPR-like proteins that might protect the upstream 3' ends. Indeed, the 3' ends of the sRNAs detected in the *atpF-atpA* and *ndhA-ndhI* intergenic regions correlated with *atpF* and *ndhA* 3' ends mapped in maize (Pfalz, et al., 2009; Zhelyazkova, et al., 2011). Additionally, 3'-RACE analysis revealed that the 3' termini of *rps16*, *ycf3*, *ndhJ* and *petD* matched the 3' ends of the sRNAs found in their 3' UTR regions (Table 8; Appendix H). The sequences corresponding to the sRNAs associated with *atpF*, *ndhA*, *ycf3* and *ndhJ* 3' ends were found highly conserved among monocots and dicots, while the *rps16* and *petD* 3' sRNAs were conserved only in monocots. (Zhelyazkova, et al., 2011). Moreover, orthologous sRNAs were detected in at least one other species (Table 8). Thus, these sRNAs also represent excellent candidates of PPR/PPR-like binding sites.

Table 8: sRNAs associated with mature 3' mRNA termini

PS	Region	Small RNA sequence[1]	Species[2]	RNA termini[3]
PS*trnK*-449	*rps16*' 3'	TATCGTGCCAATCCAACATAAGCCCCT	Hv, Os, Zm	***rps 16* 3'**(Hv)
PS*atpA*-52	*atpF-atpA*	AATTTAGGCATTATTTTTCCCCTT	Hv, Os	*atpF* 3' (Zm)
PS*psaA*-584	*ycf3* 3'	TTTGTTTTTATGTTATTTTGTGAAG	Hv, Os, At	***ycf3* 3'** (Hv)
PS*trnT*-2006	*ndhJ* 3'	AACTTTGTATCGCGCACATGACT	Hv, Os, Zm, At	***ndhJ* 3'** (Hv)
PS*trnH*-6834	*petD* 3'	ATTATTTTATTATGATCCATTTCGCG	Hv, Os	***petD* 3'** (Hv)
PS*ndhI*-49	*ndhA-ndhI*	CCAAACAAGAGAAAGAAACATAT	Hv, Os	*ndhA* 3' (Zm)

[1]T is shown instead of U in the sequences of the sRNAs. Each sRNA represents a population of molecules with ends mapping within several nucleotides of the given sequence

[2]The species in which the sRNAs have been reported are given as in Table 3.

[3]The indicated 5' and 3' RNA termini match 5' and 3' termini of the corresponding sRNA. Species in which the end has been detected are indicated in parentheses. The RNA termini verified here by 5'/3'-RACE or cRT-PCR are marked in bold (Appendix G).

The sRNA mapping downstream of *petD* is shown in Figure 22. This sRNA is found immediately adjacent to and downstream of a stable RNA stem-loop structure (dG=-21.12kcal/mol; sequence in red) whose ortholog is described to mark the processed *petD* 3' end in spinach (Chen and Stern, 1991). The 3' ends of *petD* mapped by 3'-RACE in this study correspond with the 3' ends of both the sRNA and the predicted stem-loop structure, suggesting that some of the *petD* transcripts might be stabilized at their 3' ends by a bound protein, whereas others by a stable structure. It still remains unclear why both protein- and stable RNA structure-mediated protection are used in *petD* 3' termini formation.

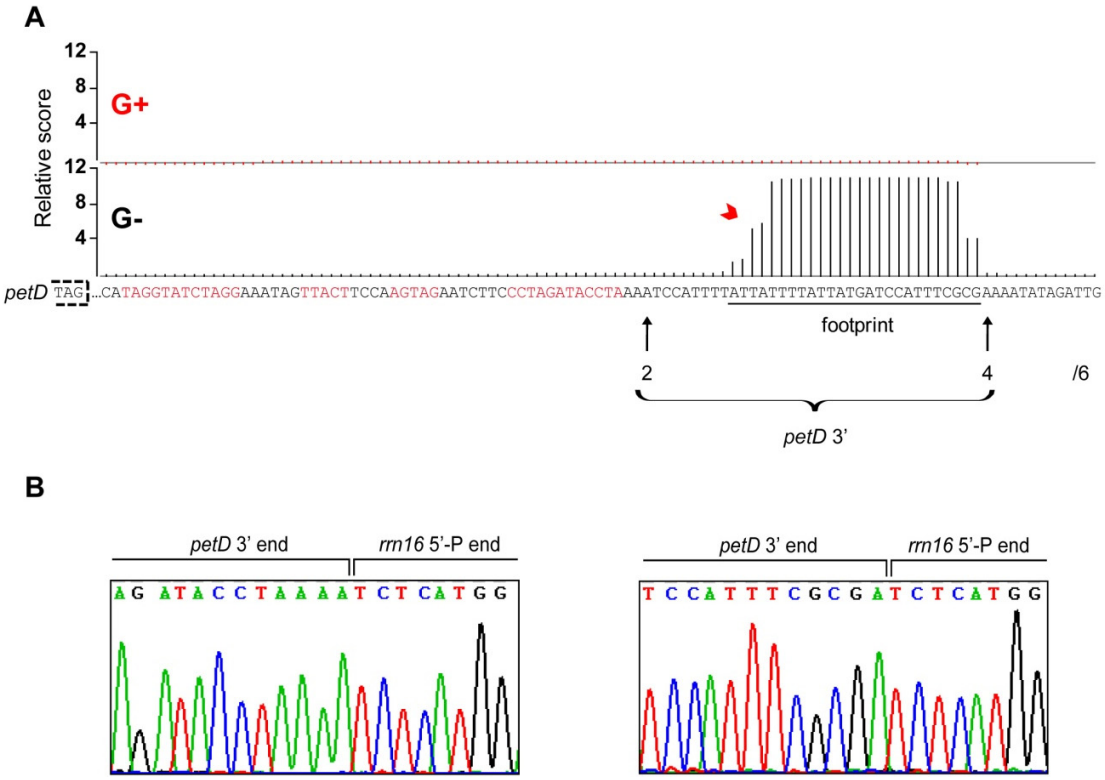


Figure 22: A sRNA associated with one of the *petD* processed 3' termini. (A) cDNA reads of G+/G- dRNA-seq libraries mapped onto the *petD* 3' UTR region. The TEX sensitive PS*trnH*-6834 (red arrow head) maps the 5' end of an unstructured sRNA associated with one of the two *petD* 3' termini mapped here by 3'-RACE analysis (marked). The other *petD* 3' end maps shortly downstream a stable stem-loop structure (sequence in red). The number of clones supporting each termini, as well as the total number of sequenced clones (number after the slash) is given. **(B)** Chromatograms displaying sequences at the ligation sites of cloned 3'-RACE products from the ligase+ reaction. The 3' ends of the *petD* and the 5'-P end of the *rrn16* which was used as a linker are shown.

In total, 14 potential mRNA 3' termini in barley chloroplasts are associated with unstructured sRNAs, representing with high probability footprints of genetically characterized or potential

PPR/PPR-like proteins, and are thus proposed to be generated *via* protein-mediated blockage of 3' nucleases (Appendix J). Eight of them correspond with previously mapped mRNA 3' termini, whereas six were experimentally verified as mRNA 3' termini in this study (Appendix H).

4 DISCUSSION

RNA-seq (massively parallel cDNA sequencing) has been shown to be a revolutionary tool for global transcriptomic analysis in both prokaryotes and eukaryotes. Here, a specialized deep sequencing approach that allows distinguishing primary from processed transcripts was used to analyze the plastid transcriptome of the monocot plant barley. The obtained genome-wide map of RNA species provided a deeper insight into chloroplast transcription and mRNA maturation. Moreover, the data obtained from this study is providing the basis for future work on various aspects of plastid gene expression regulation.

4.1 dRNA-seq is a powerful tool to catalogue primary and processed plastid RNAs

dRNA-seq is a novel deep sequencing approach designed to discriminate between primary (5'-PPP) and processed (5'-P) prokaryotic transcripts (Sharma, et al., 2010). The essence of this method is the usage of a 5'-P dependent terminator exonuclease (TEX), which degrades processed RNAs and thus enriches the sample for primary transcripts in relative terms. The method has already been successfully used to unravel the primary transcriptomes of several bacteria and archaeal species (Jäger, et al., 2009; Mitschke, et al., 2011; Sharma, et al., 2010). In this study, for the first time, dRNA-seq was used to analyze the global transcriptome of plastids. Plastids, due to their cyanobacterial ancestry, display numerous bacterial features of gene expression. As a consequence, plastid RNAs generated by transcription initiation are believed to carry a triphosphate moiety at their 5' ends, whereas RNAs generated *via* processing should have a 5'-monophosphate (Stern, et al., 2010). Thus, dRNA-seq was considered to be a suitable method for an in depth analysis of the plastid transcriptome.

An initial aim of this work was to evaluate the ability of dRNA-seq to reliably differentiate between 5'-PPP and 5'-P plastid RNAs. The following observations were made: (i) TEX treatment led to a characteristic redistribution of gene specific cDNAs between TEX+ and TEX- libraries (Figure 7). This allowed for the global mapping of transcription start sites (TSSs; primary transcripts) and processing sites (PSs; processed transcripts) in barley plastids; (ii) dRNA-seq detected with high accuracy 11 out of the 12 barley chloroplast TSSs mapped by alternative methods in previous studies (Table 3); (iii) TEX inability to efficiently degrade

processed transcripts with stable secondary structures can lead to the wrong classification of a 5'-RNA as a primary transcript (Sharma, et al., 2010). However, a computational analysis revealed that primary transcripts mapped by dRNA-seq in this study had on average a low tendency to fold into a stable structure at their 5' ends (Figure 8); (iv) 5'-RACE analysis verified 40 novel plastid TSSs identified by dRNA-seq as 5'-PPP in independent RNA preparations from barley plastids, among them candidates with predicted relatively stable structure at their 5' ends, (Appendix B); (v) Eleven processed 5' mRNA ends that were here identified by dRNA-seq were in agreement with previously published data on mature plastid 5' termini (Appendix G); (vi) 5'-RACE analysis additionally verified nine of the processed 5' mRNA ends as 5'-P transcripts in independent RNA preparations (Appendix H). Taken together, these results clearly indicate that dRNA-seq is a suitable method to differentiate between 5'-PPP and 5'-P transcripts in plastids, and can thus serve for the accurate mapping of both primary and processed plastid RNA species.

Only a small number of TSSs, mapped during the present study by 5'-RACE or previously described in other species, were not detectable as enriched transcripts in the TEX-treated library (Appendix A-1; *Not enriched*). There are several possible explanations for the observed phenomenon: (i) These 5' ends could represent identical primary and processed 5' ends, which would hinder their unequivocal identification by dRNA-seq. Indeed, identical initiation and processing sites were described for several transcripts in *Arabidopsis* mitochondria (Kühn, et al., 2005); (ii) Recently it was demonstrated that the RNA polymerase of the Gram-negative bacterium *Pseudomonas aeruginosa* may use nanoRNAs as primers for transcription initiation, thus generating primary transcripts with 5'-P ends (Goldman, et al., 2011). So far, there is no indication that such a process takes place in plastids. However, a potential analogous function of PEP, the bacteria-type RNA polymerase, would explain why some primary transcripts were not found enriched after the TEX treatment only in green libraries; and (iii) RppH, a member of the Nudix (for nucleoside diphosphates linked to some moiety X) hydrolase enzyme family, was shown to remove the beta and gamma phosphates of the initiating nucleotide in primary transcripts, and thus initiated rapid RNA decay in *E. coli* (Deana, et al., 2008). Moreover, an analogous activity was found in *Bacillus subtilis* (Richards, et al., 2011). In *Arabidopsis* there are 27 genes encoding Nudix hydrolase homologues, with seven and eight of them being targeted to mitochondria and chloroplasts, respectively. Organellar-type Nudix hydrolases showed pyrophosphohydrolase activity towards various substrates, such as ADP-ribose, ADP-glucose,

coenzyme A, and NADH (Ogawa, et al., 2008). However, it still remains to be elucidated if one or more of these enzymes also exhibit RppH-like activity, and thus generate monophosphorylated 5' ends of organellar primary transcripts.

4.2 The primary transcriptome of barley *albostrians* plastids

In this work, dRNA-seq was used to study the primary transcriptome of green and white plastids of the barley *albostrians* mutant line. Green *albostrians* plastids are phenotypically identical to wild-type chloroplasts, and thus their transcription relies on both PEP and NEP activity. Here, they will be also referred to as barley chloroplasts. White plastids are ribosome deficient, and therefore not able to synthesize plastid-encoded proteins, among them the core subunits of PEP. As a consequence, transcription in these plastids is solely driven by the nuclear encoded polymerase NEP (Hess, et al., 1993). Cataloguing the primary RNAs in barley *albostrians* plastids not only revealed general features of chloroplast transcription, but also allowed for the analysis of the division of labor between PEP and NEP in plastid gene expression. Furthermore, the most comprehensive up to date list of TSSs in mature chloroplasts and PEP-deficient plastids generated in this study gave the opportunity to address the question of conserved PEP and NEP promoter motifs in the monocot plant barley. Last but not least, this data revealed the existence of numerous ncRNA generated *via* transcription of free-standing genes.

4.2.1 General features of chloroplast transcription revealed by dRNA-seq

The current view on transcription in chloroplasts is mainly derived from the analysis of a few individual genes and RNAs. Therefore, it remains unclear if the observed features are rules rather than exceptions. Here, dRNA-seq analysis was used to obtain a genome-wide map of the transcription start sites (TSSs) in barley chloroplasts. In total, 176 TSSs were discovered in green *albostrians* plastids, and thus this work presents the most comprehensive up to date list of chloroplast transcription initiation sites. The obtained data revealed the following general features of chloroplasts transcription and transcriptional organization (Figure 23): (i) The majority of chloroplasts TSSs are involved in the synthesis of annotated genes; (ii) Internal promoters (gTSSs) are detectable in most plastid operons. Unlike in bacteria, polycistronic mRNAs are usually processed in smaller polycistronic or monocistronic transcripts in plastids. If

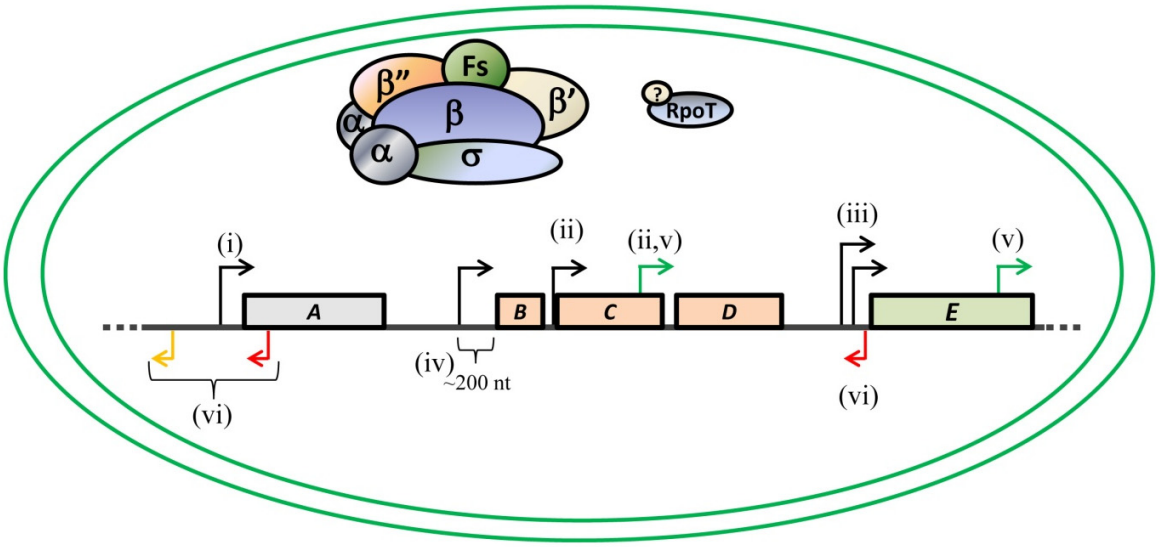


Figure 23: General features of transcription in mature barley chloroplasts revealed by dRNA-seq analysis. The figure provides a schematic representation of the general features of chloroplast transcription (numbered in accordance to the description in chapter 4.2.1) by depicting the TSS distribution in the barley plastome. TSSs are indicated by arrows and color coded based on their annotation: gTSS (black), iTSS (green), aTSS (red) and oTSS (orange). Genes are represented by squares. Genes with the same colors form an operons, i.e. *B-C-D*. PEP and NEP are shown, with PEP being enlarged since it is the major polymerase in mature chloroplasts.

this intercistronic mRNA processing has evolved to facilitate translation by plastid ribosomes, is still a matter of debate (Barkan, 2011). The presence of numerous operon-internal TSSs that was observed in this study is an alternative way to adjust the abundance of individual mRNAs encoded in the same operon; (iii) Multiple promoters drive transcription of many plastids genes. Alternative promoters allow for the differential regulation of genes under distinct conditions, as previously shown for the barley *psbD-psbC* (Sexton, et al., 1990). In the case of protein encoding genes, the usage of multiple promoters would in addition result in the generation of mRNAs with different lengths of 5'UTRs from the same gene/operon. If this is a mechanism to control plastid mRNA stability and/or translation efficiency remains to be still elucidated (cf. Cahoon, et al., 2004). Moreover, in analogy to complex eukaryotes, the usage of multiple promoters could allow for the ubiquitous or tissue-specific expression of plastid genes (Ayoubi and Van De Ven, 1996); (iv) Primary mRNAs in chloroplasts carry long 5'-untranslated regions. This is in good agreement with the fact that 5'-UTRs are important sites of posttranscriptional regulation in plastids (Barkan, 2011; Stern, et al., 2010). Thus, binding of more than one protein to this region might be a general phenomenon. Alternatively, such long 5'UTRs could indicate the necessity

for complex folding of this region, which might be a requirement for riboswitch formation or recognition and binding of certain proteins; (v) Similar to bacteria, chloroplasts utilize several TSSs within annotated genes (Mitschke, et al., 2011; Sharma, et al., 2010). The function of iTSSs that do not participate in the expression of downstream genes remains unclear. Several iTSSs in cyanobacteria were speculated to produce shorter isoforms of proteins (Mitschke, et al., 2011). Alternatively, some gene-internal transcripts could serve as target mimicry for ncRNAs, as already shown for plant miRNAs (Franco-Zorrilla, et al., 2007). (vi) TSSs in intergenic regions and opposite to annotated genes lead to the synthesis of numerous chloroplast ncRNAs (further discussed in Chapter 4.2.3.).

4.2.2 Division of labor between PEP and NEP

Many efforts have already been directed to shed light on the function and complex interplay of PEP and NEP in plastid transcription. However, the current notion of the division of labor between the two plastid polymerases remains rather unclear (Liere, et al., 2011). In this study, a genome-wide analysis of the distribution of transcripts synthesized by these two enzymes was used as a means to learn more about the role of PEP and NEP in plastid transcription. Indeed, the analysis and comparison of the transcription initiation sites generated in green (transcription by PEP and NEP) and white (transcription by NEP) barley *albostrians* plastids revealed several novel, as well as previously observed aspects of PEP and NEP-dependent transcription in plastids.

The PEP-deficient white *albostrians* plastids were included in this analysis in order to allow for the unambiguous identification of NEP promoters. Indeed, numerous plastid transcription initiation sites were detected in white tissue, even more than the TSSs mapped in the phenotypically wild-type green plastids. The observation of active transcription in the absence of the plastid encoded RNAP is in agreement with previous studies in transplastomic tobacco plants with abolished PEP activity (Krause, et al., 2000; Legen, et al., 2002). Still, the high number of promoters used in white *albostrians* leaves was rather surprising. In tobacco, large transcripts were shown to accumulate in plastids lacking the PEP genes *rpoA*, *rpoB* or *rpoC1* (Krause, et al., 2000). TEX is not capable to efficiently degrade the 5'-structured RNAs that might have been generated by processing or degradation of such high molecular weight transcripts (potentially generated from only a few NEP promoters), and thus they might have been misannotated as

primary transcripts in this study. However, a computational analysis revealed that stable structures are unlikely to be formed near the 5' ends of most analyzed transcripts in white plastids. In addition, no evidence was found for the occurrence of dsRNA species (theoretically present in RNA-seq libraries by cDNA reads precisely/highly complementary to each other) which are often protected from exonucleolytic attack and could also lead to a wrong assignment of TSSs. Moreover, 25 of the 30 TSSs selected for further analysis by 5'-RACE were successfully verified in independent RNA preparations from white material. A possible reason for the higher number of TSSs mapped by dRNA-seq in white compared to green plastids could be a contamination of white RNA preparations with non plastid-encoded transcripts. The nuclear and mitochondrial genomes of higher plants contain DNA sequences of chloroplast origin that still may show a high degree of similarity or even identity to the original sequence (Hao and Palmer, 2009; Noutsos, et al., 2005). Though usually not active, some might be transcribed and contained in the plastid samples used in this study, in particular in those from the less pure plastids isolated from white leaves. Such 'chloroplast-like' transcripts from the nucleus would be expected to carry non-chloroplast sequences at the 5' end, permitting their detection. In the case of 'chloroplast-like' transcripts of mitochondrial origin, the situation is more complicated, since the mitochondrial RNA polymerases recognize NEP promoters (Liere, et al., 2011). Thus, sequences that are detected in samples from plastid preparations, but are homologous in both organelles, might have originated from mitochondria. The lack of a complete mitochondrial genome sequence of barley, however, precludes the analysis of such mitochondria-borne sequences. Arguably, such RNAs should be detected in this study, because with increasing length they should become polymorphic or chimeric, i.e. they should have acquired SNPs, or might be fused to non-chloroplast sequences at the 3' end, respectively. So far, no chimeric sequences of any type were found in white (or green) libraries. Moreover, in general, mitochondrial transcripts are of much lower abundance than chloroplasts RNAs, thus further minimizing the danger of a potential contamination of samples from white tissue with transcripts of mitochondrial origin. Taken together, the above listed observations and arguments indicate that the numerous TEX-resistant 5' ends mapped in white leaves are indeed generated *via* transcription from plastid NEP promoters.

The comparison of the TSSs mapped in green (both PEP and NEP present) and white (only NEP present) barley *albostrians* plastids was used to study RNAP specific gene expression in

mature chloroplasts. NEP activity was reported to be higher in white compared to green leaves (Emanuel, et al., 2004), and therefore plastids of green leaves are unlikely to exhibit the activities of NEP promoters that are not detectable in white tissue. Thus, the intersection of the TSS populations detected in green and white dRNA-seq libraries should indicate NEP-dependent primary transcripts occurring in mature chloroplasts. In total 176 and 244 TSSs were detected in green and white dRNA-seq libraries, with only 22 TSSs found to be identical in both plastid types (Figure 9B). Therefore, the high-throughput data obtained in this study clearly indicates that PEP is the dominating RNA polymerase in green leaves, synthesizing at least 88% of the detected transcripts, while just a minority of RNAs in mature chloroplasts originates from NEP activity. This observation is in good agreement with previous studies on chloroplasts transcription (Demarsy, et al., 2006; Zoschke, et al., 2007) and can be attributed to the higher usage of NEP promoters in early compared to late leaf development and increasing transcriptional activity of PEP during chloroplast maturation (Baumgartner, et al., 1993; Courtois, et al., 2007; Demarsy, et al., 2011; Emanuel, et al., 2004; Kapoor, et al., 1997; Swiatecka-Hagenbruch, et al., 2008; Zoschke, et al., 2007). According to previously published data the following genes/operons are transcribed by NEP in green leaves: *rpoB*, *rpl23* and *rps15* (Hübschmann and Börner, 1998; Liere and Maliga, 1999; Swiatecka-Hagenbruch, et al., 2007). Here, the corresponding TSSs T*rpoB*-147, T*rpl23*-71 and T*rps15*-228 were found in both green and white *albostrians* plastids. During the current study it became clear that PEP and NEP can obviously generate transcripts with identical 5' ends (T*psbA*-80, T*psaA*-209, and T*psbN*-43, see Chapter 3.2.6.4.). Therefore, it remains unclear which polymerase, NEP or PEP, is responsible in mature chloroplasts for the usage of several here newly identified TSSs that are identical in both plastids types. Consequently, this work can so far list the following genes and ncRNAs only as *potentially* transcribed via a NEP promoter in chloroplasts: *trnK* (T*trnK*-239), psbM (T*psbM*-348), *rps2* (T*rps*2-152), *ndhC* (T*ndhC*-336), *trnP* (T*trnP*-21), *rps8* (T*rps8*-142); *ndhB* (T*ndhB*-275), *ndhI* (T*ndhI*-99), *rpoC1* (T*rpoC1*-599), Hv_nc3 (T*psbK*-783), Hv_nc39 (as_*rps18*; T*trnP*-1937), Hv_nc46 (as_*petD*; T*psbN*-3371), Hv_nc56 (as_*rps15;* T*trnN*-1479) and Hv_nc58 (as_*ndhF*; T*rpl32*-1224). More detailed studies will be necessary to clearly assign a polymerase to the above listed examples. For example, *in vitro* transcription experiments might be used to determine if PEP or NEP (or both) are active on the isolated promoters. Additionally, polymerase-promoter interactions can be confirmed by electrophoretic mobility shift assays and

the exact polymerase binding site can be subsequently mapped by DNase footprints. Alternatively, if conserved, the occurrence of the above listed TSSs can be analyzed in transgenic *Arabidopsis* lines with altered or abolished RpoTp or RpoTmp activity (Swiatecka-Hagenbruch, et al., 2008). In addition, comparing the abundance of these transcripts in young and mature chloroplasts could also give additional hints for PEP or NEP-dependent transcription.

Altogether, 244 TSSs were detected in white plastids, among them 128 gTTSs upstream of 70 plastid genes. Many of the NEP-transcribed genes encode proteins of the photosynthetic apparatus, which until now were thought to be expressed solely by PEP (Hajdukiewicz, et al., 1997; Swiatecka-Hagenbruch, et al., 2007). Transcription of all genes was previously observed in Δ*rpoB* (PEP-deficient) tobacco plastids and mainly attributed to spurious transcription initiation by NEP throughout the plastome (Legen, et al., 2002). However, the high-resolution TSS mapping analysis in this study revealed that RNAs in white plastids are not transcribed from randomly distributed AT-rich sequences or YRTa motifs, but rather result from distinct initiation sites localized directly upstream of the genes. Moreover, NEP was observed to generate virtually the same mRNAs as PEP, with similar 5'UTR and 3'UTR lengths and, in the case of *psaA*, *psbA*, and *psbN,* even the same 5' primary ends. Interestingly, also other general features of transcription and transcriptional organization in PEP-deficient white plastids were found to be very similar to the ones reported for wild type chloroplasts (see Chapter 4.2.1.). This observation indicates that the overall distribution of NEP promoters is similar to the one of PEP promoters and that in spite of the different nature of the RNA polymerases, both PEP and NEP might still perform similar functions in plastid transcription. Therefore, it is unclear why NEP cannot rescue the albino phenotype in Δ*rpo* tobacco plastids, *i.e.,* PEP-deficient mutants with functional translation. A possible explanation might be that the amount of transcripts generated by NEP instead of PEP might be simply insufficient to meet the demands for the extensive translation required for greening (Allison, et al., 1996; Siemenroth, et al., 1981). Alternatively or additionally, NEP might not transcribe a small number of genes, among them several *trn* and the photosynthetic genes *psbE-F-L-J* and *petN*, for which TSSs were found in green but not in white plastids. Furthermore, mRNA stability and even translational efficiency may be varying with differential RNAP usage (Cahoon, et al., 2004; Legen, et al., 2002).

In *Arabidopsis*, the activation of a NEP promoter was described to compensate for abolished transcription from the plastid *atpB* PEP promoter (Schweer, et al., 2006). Here, we observed that

in the absence of PEP at least 222 NEP promoters are utilized in white tissue that could not be detected in green plastids. Therefore, activation of transcription by NEP in the absence of PEP might serve as a general rescue mechanism in higher plants. The observed stimulation of NEP-dependent transcription might also be a cause or consequence of the higher accumulation of *RPOTp* transcripts in white *albostrians* leaves (Emanuel, et al., 2004). Additionally or alternatively, PEP, being the more abundant and active RNA polymerase in mature chloroplasts, might hinder the access of NEP to its promoters, if they are located in close spatial proximity to PEP promoters (Lyubetsky, et al., 2011). Indeed, many of the NEP promoters detected in this study were found in close proximity to PEP promoters. Therefore, it would be interesting to experimentally determine the affinities of NEP and PEP to their promoters. In addition, *in vitro* competition experiments could reveal to what degree NEP transcription is stimulated if PEP amounts are reduced, but not fully lacking.

The question still remains as to whether the numerous NEP promoters found in white mutant plastids are also active in normal wild-type plastids. The *Arabidopsis RPOTmp* gene coding for a mitochondrial and chloroplast (NEP) RNA polymerase is particularly active in non-green tissues and in very young leaves (Emanuel, et al., 2006). Moreover, the expression of the previously described NEP-dependent genes was distinctly higher in developing compared to mature chloroplasts (Courtois, et al., 2007, Zoschke, et al., 2007). Similarly, NEP-dependent transcripts were shown to contribute substantially to the plastid RNA population during stratification and germination of *Arabidopsis* but not afterward (Demarsy, et al., 2011). Therefore, it should be worth investigating non-green seed proplastids and developing chloroplasts for the usage of the NEP promoters that are exclusively detected in white *albostrians* plastids.

4.2.3 Promoter motifs in green and white *albostrians* plastids

The current view on PEP and NEP consensus promoter elements is rather speculative due to the fact that until now only a few plastid TSSs have been mapped and their promoters analyzed (Liere, et al., 2011). A main reason for this data limitation is that commonly used methods for TSS mapping, such as primer extension, in vitro capping by guanyltransferase, and 5'-RACE, target only individual transcripts (Hübschmann and Börner, 1998; Liere and Maliga, 1999; Swiatecka-Hagenbruch, et al., 2007). In this work, dRNA-seq provided the means for a high throughput mapping of plastid TSSs on a genome wide scale. Thus, the analysis of the upstream

sequences of the TSSs mapped in green and white *albostrians* plastids revealed prevalent PEP and NEP promoter motifs in the monocot plant barley.

In green plastids, both PEP and NEP primary transcripts exist, with the former being the predominant RNA polymerase generating at least 88% of all primary transcripts. Consequently, MEME analysis detected the typical motifs of PEP promoters as the only conserved sequences in the upstream region of the TSSs mapped in green plastids. However, the -10 and -35 boxes were discovered upstream of only 25% and 11% of all TSSs, respectively. The main reason for the poor detection by MEME was attributed to the high variability of the nucleotides in the PEP promoter elements, since a manual search with relaxed parameters increased their detection rate: the -10 box was mapped in 89% of the TSSs, while the -35 box in 70% of the TSSs with a -10 element. Still, the lower conservation of the -35 box indicates that PEP transcription can occur independently of the -35 element. Additionally, PEP may utilize motifs with extremely low similarity to the *E. coli* σ^{70} -35 element or other variable *cis* elements.

All TSSs in white plastids are generated by NEP activity. However, the YRTa motif typical for NEP Type-Ia and -Ib promoters (Liere, et al., 2011) was detected upstream of only 73% of the TSSs in white leaves. Similarly, no conserved promoter motifs could be detected upstream of many transcription start sites in higher plant mitochondria, where transcription is performed by NEP-related phage-type RNA polymerases (Liere, et al., 2011). In addition, no evidence was found for a conserved GAA-box in a regular distance to the mapped YRTa motifs as described for so-called Type-Ib NEP promoters in dicots (Liere and Börner, 2007). Transcription from tobacco P*clpP*-53, the only example of a NEP Type-II promoter, was shown to be dependent also on promoter elements located downstream of the TSSs (-5 to +25; Sriraman, et al., 1998). MEME did not detect any conserved downstream motif when the analyzed region around the TSSs in white leaves was extended to -50 to +25. Thus, Type-Ia was the only NEP promoter type clearly identifiable in barley plastids. If this is related to the fact that barley and other grasses have only one NEP polymerase (RPOTp) in contrast to eudicots, which have two NEP polymerases (RPOTp and RPOTmp; Liere, et al., 2011), remains to be further analyzed.

4.2.4 Non-coding RNAs in plastids

Over the last years, non-coding RNAs (ncRNAs) have emerged as central players in both prokaryotic and eukaryotic regulation of gene expression (Prasanth and Spector, 2007; Repoila

and Darfeuille, 2009). In bacteria, ncRNAs are the most predominant class of post-transcriptional regulators (Papenfort and Vogel, 2009). Recently, a high degree of non-coding RNA synthesis was also detected in cyanobacteria, the chloroplast progenitors (Georg, et al., 2009; Mitschke, et al., 2011). In addition, several cyanobacterial ncRNAs were shown to participate in the control of various cellular processes (Georg and Hess, 2011). Until recently, only a few reports supported the existence of regulatory ncRNAs in chloroplasts. These studies focused on single RNAs and did not point to ncRNA synthesis as a more general regulatory mechanism in chloroplasts (Georg, et al., 2010; Hotto, et al., 2010; Nishimura, et al., 2004; Sharwood, et al., 2011; Zghidi-Abouzid, et al., 2011). In this work, dRNA-seq analysis of the transcriptome of barley chloroplasts provided evidence for the extensive ncRNA synthesis in plastids. The observation that ncRNAs are common in chloroplasts is in agreement with the outcome of three other RNA-seq studies carried out in parallel to this work (Hotto, et al., 2011; Mohorianu, et al., 2011; Wang, et al., 2011). Yet, due to the experimental setup, these studies did not directly detect ncRNAs generated via transcription (rather than processing or degradation) - the so far most abundant class of known regulatory ncRNAs in bacteria (Repoila and Darfeuille, 2009). In contrast, the TEX-based RNA-seq approach used here allowed for the unequivocal identification of ncRNAs existing as primary transcripts in chloroplasts. Moreover, the genome-wide mapping of 5'-P chloroplasts transcripts in this analysis revealed no indication of abundant processed transcripts antisense to genes or in intergenic regions (see Chapter 3.3.)

Numerous TSSs in intergenic regions and opposite to annotated genes were mapped in both green and white *albostrians* plastids (Figure 9; see Chapter 3.2.3. and 3.2.7.), demonstrating the existence of non-coding RNAs in both plastid types. The set of aTSSs and oTSSs in green plastids let to the prediction of 60 ncRNAs in mature barley chloroplasts. The TSSs of eight selected candidates were additionally verified by 5'-RACE analysis. Moreover, 3'-RACE analysis determined the 3' ends of seven of them, and thus confirmed the synthesis of ncRNAs from the corresponding TSSs. Eleven ncRNA candidates were found to be transcribed from adjacent promoters in both green and white *albostrians* plastids, i.e. have both PEP and NEP promoters. Moreover, the TSSs of Hv_nc3 (T*psbK*-783), Hv_nc39 (as_*rps18*; T*trnP*-1937), Hv_nc46 (as_*petD*; T*psbN*-3371), Hv_nc56 (as_*rps15;* T*trnN*-1479) and Hv_nc58 (as_*ndhF*; T*rpl32*-1224) were detected to be identical in both plastid types, and therefore these ncRNAs could be transcribed via potential NEP promoters in chloroplasts (Appendix E). NEP-dependent

ncRNAs were also detected in mature *Arabidopsis* chloroplasts (Hotto, et al., 2011). In theory, non-coding RNA synthesis via PEP and/or NEP promoters could allow for differential composition and abundance of ncRNAs in tissues or throughout chloroplast development.

Based on their genomic position, the majority (48) of the ncRNAs detected in this study can be classified as potential *cis*-encoded ncRNAs. Antisense RNAs were detected to ~35% of all genes in barley chloroplasts (Table 5). Both photosynthesis and genetic system genes were found associated with at least one candidate. Interestingly, antisense transcripts were detected to coding as well as non-coding regions. Thus, similar to the complex scenario found in bacteria (Brantl, 2007; Repoila and Darfeuille, 2009), *cis*-encoded asRNAs could potentially regulate gene expression in plastids on multiple levels and via diverse mechanisms. For example, eight ncRNAs (e.g., as*_atpI*, as*_rps15,* as*_psaA)* were found complementary to the 5'-UTR region, and thus may function to repress transcription or translation, as in bacteria (Brantl, 2002). Another ten asRNAs can hypothetically basepair with mRNA 3' UTRs, and thus participate in 3' end formation by blocking 3' to 5' exonucleases. Such a mechanism is already described in *E.coli* (Opdyke, et al., 2004) and an asRNA against *psbT* in *Arabidopsis*, reported to stabilize the complementary mRNA, could function in a similar manner (Zghidi-Abouzid, et al., 2011). ncRNAs were also detected against the 5' and 3' regions of tRNA and rRNA precursors and might potentially interfere with their proper maturation. The majority of *cis*-encoded asRNA candidates detected in this study were found complementary to annotated ORFs. An internal asRNA was shown to regulate the expression of a photosynthesis gene in cyanobacteria by forming asRNA-mRNA duplex which was targeted for degradation (Duhring, et al., 2006). Given the cyanobacterial ancestry/origin of plastids, similar ncRNA-dependent mechanism for adjusting mRNA levels may also operate in chloroplasts. Last, but not least, twelve ncRNAs were found against introns of plastid genes, and thus might be involved in regulating RNA splicing and maturation of the corresponding precursor transcripts. Indeed, as*_petB*(1), as*_trnV* and as*_rpl2* were found to be complementary to essential regions involved in base-pairing interactions required for splicing of group II introns in plastids (Michel and Ferat, 1995).

A recent study discovered 107 *cis*-encoded asRNA candidates in *Arabidopsis* using RNA-seq (Hotto, et al., 2011). The comparison of the potential *cis*-encoded asRNAs discovered in barley and *Arabidopsis* revealed that there is just a small overlap between the two lists. For example, even though not conserved in position or sequence, several ncRNAs were found to target similar

coding or non-coding regions of chloroplast genes, i.e. in both barley and *Arabidopsis* ncRNAs were found antisense to *psbD*, *rpoC2*, *atpA* and *ndhD* ORFs, *psbL-psbF* intergenic region, 3' of *trnL* precursor, and *petB* and *trnV* introns. Moreover, Hv_nc22 (asRNA against *psaB-rps14* intergenic region), Hv_nc48 (asRNA against *rpl2* intron) and Hv_nc34 (asRNA against *rbcL* 3' UTR) were found partially overlapping with *Arabidopsis* ncRNAs from the corresponding regions but still not sharing the same 5' ends (data not shown). A possible explanation for the low correlation between the *cis*-encoded asRNAs discovered in barley and *Arabidopsis* could be that plastid ncRNAs are poorly or not at all conserved among monocots and dicots. Alternatively, the expression of ncRNAs might be strongly dependent on growth conditions and developmental stage. Thus, the generated lists of ncRNAs in barley and *Arabidopsis* chloroplasts might be simply too heterogeneous to allow for a comparative analysis between these species. In order to resolve this issue, it would be worth investigating the conservation of *cis*-encoded asRNAs in the chloroplast transcriptome of a monocot and a dicot plant using the same experimental approach and conditions. Still the possibility remains that the low conservation of asRNAs between barley and *Arabidopsis* might also indicate that a large part or even all of these RNAs actually do not have a function in gene expression but rather represent "transcriptional noise".

The majority of bacterial regulatory ncRNAs described so far function as *trans*-encoded asRNAs, i.e. act on one or more elsewhere encoded RNAs through short regions of complementarity (Storz, et al., 2005). Base pairing of the ncRNA to sequences adjacent to/overlapping with the ribosomal binding site (RBS) region and/or the start codon of an mRNA target can lead to translation inhibition. Alternatively but less common, pairing of the ncRNA with the 5'-UTR region shortly upstream the translation initiation region can inhibit the formation of a structure that sequesters the RBS, and thus results in translation activation (Repoila and Darfeuille, 2009; Vogel and Wagner, 2007). In this study, twelve ncRNA candidates were found in intergenic regions and it could be therefore speculated that they can exhibit a potential activity as *trans*-encoded asRNAs in barley chloroplasts. There are already several reports suggesting that *cis*-encoded asRNAs function in chloroplast gene expression regulation (Georg, et al., 2010; Hegeman, et al., 2005; Sharwood, et al., 2011; Zghidi-Abouzid, et al., 2011). In contrast, the question of *trans*-encoded asRNA activity in plastids has not yet been addressed. Therefore, it was interesting to examine if these ncRNAs could participate in

interactions similar to bacteria, i.e. leading to translation inhibition. IntaRNA, a computational analysis already proven to effectively predict ncRNA targets (Busch, et al., 2008; Richter, et al., 2009), was used to screen the sequences around the start codon of barley plastid mRNAs for short regions of nearly (perfect) sequence complementarity, i.e seed regions to the ncRNA candidates. Indeed, numerous potential mRNA-ncRNA interactions were revealed and it will be worth further investigating the best scoring candidates (Appendix F). For example, secondary structure computational analysis could be used to predict if the seed regions are positioned in a single stranded region of the ncRNA, and thus are available for pairing with the target mRNA sequences. Afterwards, selected interaction partners could be experimentally tested in *E.coli*. This might be done by using for example reporter gene expression assays in which a target-reporter gene fusion can reveal if the presence of a certain ncRNA leads to its translation inhibition (Richter, et al., 2009). However, the best method for studying function and relevance of *trans*-encoded, as well as *cis*-encoded ncRNA in plastids remains the analysis of transplastomic plants lacking or over-expressing plastid ncRNAs. As up to date no routine protocol for the genetic manipulation of barley plastids exists, it will be very useful to reveal the ncRNA composition in tobacco chloroplasts by dRNA-seq in order to be able to address this important question *in vivo*.

4.3 The processed transcriptome of barley chloroplasts

The combination of dRNA-seq with 3'-RACE allowed us to globally map both 5' and 3' processed mRNA ends, as well as associated sRNA representing footprints of PPR/PPR-like proteins in mature barley chloroplasts. The obtained data provides evidence for the overall impact of protein-mediated protection from ribonucleases on processed mRNA termini formation. Furthermore, several 5' and 3' mRNA termini mapped in this study support a model suggesting that intercistronic processing involves site-specific blockage of exonucleases. In addition, the positional bias of PPR footprints in 5'-UTRs suggests that PPR proteins might play important roles in both RNA stability and translational activation.

4.3.1 Protein-mediated protection - the predominant mechanism for processed mRNA termini formation in chloroplasts

Numerous PPR and PPR-like proteins have been reported to be involved in plastid RNA processing and stabilization (Barkan, 2011; Stern, et al., 2010). In maize, PPR10 was demonstrated to specifically bind to two intercistronic mRNA regions and act as a protein barrier to 5' and 3' exonucleases, thus resulting in the accumulation of processed RNAs (originating from different precursor molecules) with 5' or 3' ends defined by the bound protein (Pfalz, et al., 2009; Prikryl, et al., 2011). It was proposed that eventually these PPR10 stabilized transcripts would be degraded by ribonucleases with the exception of the region protected by the bound protein (Pfalz, et al., 2009). Indeed, such potential PPR10 footprints were detected as small RNAs (sRNAs) in the transcriptomes of rice and maize (Johnson, et al., 2007; Morin, et al., 2008; Pfalz, et al., 2009).

In this work we investigated the occurrence of sRNAs representing footprints of bound proteins in the transcriptome of mature barley chloroplasts. This analysis was done in collaboration with A. Barkan. The following observations and conclusions were made: (i) Processed 5' mRNA ends in barley were found associated with abundant, small RNAs corresponding to previously determined binding sites of the PPR proteins PPR10, HCF152 and PGR3 (see Chapter 3.3.2.1.; Cai, et al., 2011; Pfalz, et al., 2009; Zhelyazkova, et al., 2011). These sRNAs matched 5' and 3' ends of processed transcripts that were shown to fail to accumulate in the absence of the corresponding proteins (Meierhoff, et al., 2003; Pfalz, et al., 2009; Yamazaki, et al., 2004). Furthermore, maize and *Arabidopsis* orthologs of the sRNAs corresponding to PPR10, HCF152, and the putative CRP1binding sites were found missing in the corresponding *ppr* mutant backgrounds (Ruwe and Schmitz-Linneweber, 2011; Zhelyazkova, et al., 2011). Taken together, these observations indicate that these small RNAs are in fact protein binding sites and thus protected from RNase-mediated degradation, i.e. they are *in vivo* footprints of PPR proteins. (ii) This analysis detected sRNAs that correspond to processed mRNA ends that were shown to require the PPR/PPR-like proteins MRL1, PPR38, HCF107, CRP1 and CRP2 for their accumulation, but for which no exact protein binding site has been yet determined (see Chapter 3.3.2.1; Barkan, et al., 1994; Fisk, et al., 1999; Hashimoto, et al., 2003; Hattori, et al., 2007; Hattori and Sugita, 2009; Johnson, et al., 2010; Sane, et al., 2005; Zhelyazkova, et al., 2011). Therefore, it is highly likely that these sRNAs represent the *in vivo* binding sites of the

corresponding proteins. (iii) sRNAs were found identical to processed 5' mRNA ends for which no stabilizing proteins has been reported so far. These sRNAs most likely also represent binding sites of unknown PPR-like proteins (see Chapter 3.3.2.2.). (iv) The 3' ends of sRNAs accumulating in close spatial proximity to upstream ORFs were shown to match the 3' termini of the mRNA of these genes (see Chapter 3.3.3.2.). This observation implies that these 3' mRNA ends might be generated *via* protection by a bound protein from 3' directional degradation.

The aforementioned sRNAs associated with mature mRNA termini mapped in this study were detected as plateaus of ~25 nt long cDNA reads in TEX- dRNA-seq library. sRNAs were predicted to lack stable secondary structure (Zhelyazkova, et al., 2011) which could hypothetically shield them from degradation (Barkan, 2011; Stern, et al., 2010). Still, they are obviously accumulating to higher levels than the mRNAs whose processed ends they define *in vivo*. Therefore, it seems very likely that the observed resistance of these small RNAs to ribonucleases-dependent degradation is indeed resulting from a bound protein. Due to several lines of evidence, we believe that the observed footprints result from RNA protection by PPR/PPR-like proteins. PPR10, for example, was already shown to bind with high affinity to an unusually long RNA stretch and thus provides an effective barrier to exonucleases. An exceptionally long and stable protein/RNA interface was proposed to be a typical feature of PPR/PPR-like proteins (Prikryl, et al., 2011). Moreover, many sRNAs were found highly conserved in angiosperms (Ruwe and Schmitz-Linneweber, 2011; Zhelyazkova, et al., 2011). This is in good agreement with the observed high amino acid sequence conservation of PPR proteins in comparison to other plant protein families (O'Toole, et al., 2008). Therefore, it is highly likely that PPR/PPR-like proteins are primarily mediating the observed protection of sRNAs, and therefore play a major role in mRNA termini formation in chloroplasts.

In total, 19 out of the 22 processed 5' mRNA ends identified in this study were found associated with sRNAs standing for excellent candidates or representing *in vivo* footprints of PPR/PPR-like bound proteins. Thus, the majority of processed 5' mRNA ends in barley chloroplasts are with a high probability generated via a specific binding of a PPR/PPR-like protein which serves as a barrier for 5' exonuclease dependent degradation (Figure 24). Concerning 3´end processing, protein-, as well stable RNA structure-mediated blockage of 3' nucleases were proposed to generate the 3' termini of 14 and 13 barley mRNAs, respectively. Moreover, many of the detected 5' and 3' mRNA ends, as well as the sRNAs associated with

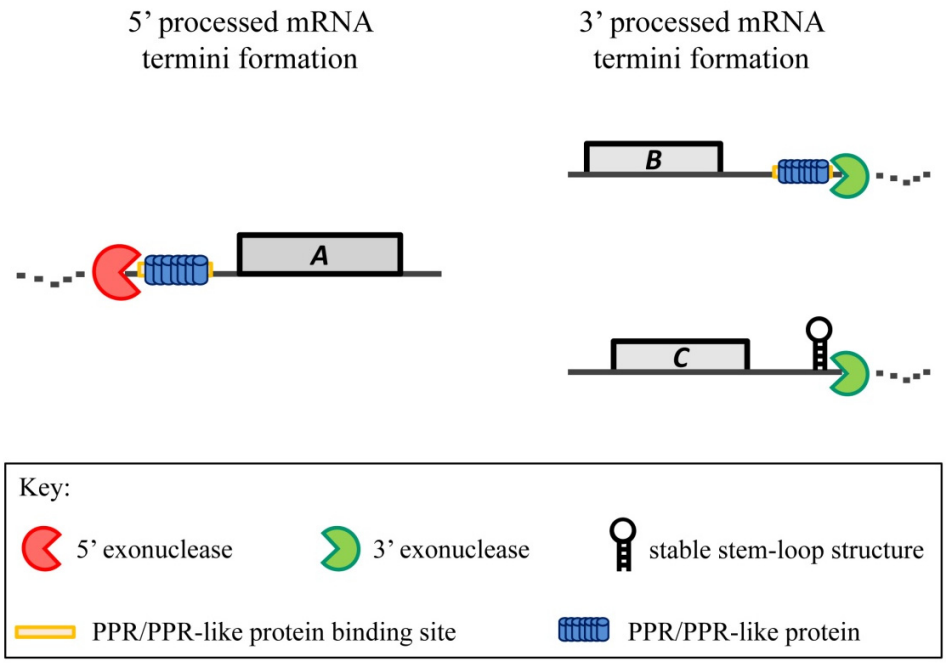


Figure 24: Mechanisms of 5' and 3' processed mRNA termini formation in barley chloroplasts. The majority (19 out of 22) of processed 5' mRNA ends identified in this study are with a high probability generated via a specific binding of a PPR/PPR-like protein which serves as a barrier for 5' exonuclease dependent degradation. 14 and 13 barley mRNA 3'ends mapped here are most likely generated *via* a PPR/PPR-like protein- and stable RNA structure-mediated blockage of 3' nucleases, respectively. RNA-structure mediated blockage is predominant at the 3' ends of monocistronic or polycistronic transcripts, whereas protein-mediated protection is more often observed at intercistronic regions

them are found conserved among monocots and dicots (Ruwe and Schmitz-Linneweber, 2011; Zhelyazkova, et al., 2011). Taken together, these observations indicate that protein-mediated protection serves as the predominant mechanism in processed mRNA termini formation in angiosperm chloroplasts.

There are more than 100 P-type PPR proteins predicted for angiosperm chloroplasts (O'Toole, et al., 2008) but only a few have been characterized until now. This study provides the basis for future work on relating novel PPR-like proteins to their specific targets of action. In total, 16 putative PPR/PPR-like footprints were found to be associated with mRNA ends for which no stabilizing protein have been yet identified. The mapping of these potential binding sites is anticipated to facilitate the characterization of the corresponding proteins and elucidation of their functions in chloroplast RNA processing.

4.3.2 General model for intercistronic mRNA processing in chloroplasts

It was initially proposed that intercistronic processing occurs *via* site-specific cleavages by endoribonucleases resulting in the formation of adjacent processed 5' and 3' termini. This model was based on low resolution mapping data that suggested that 5' and 3' processed ends are spatially positioned next to each other in several intercistronic regions (Barkan, 2011). Moreover, this hypothesis was favored over the idea of exonucleases dependent mRNA processing since at that time no bacterial 5' to 3' exonucleases were known, while endoribonucleases activity was well documented in both bacteria and chlorolasts (Stern, et al., 2010). Recently, RNase J, initially believed to act only as an endonuclease, was further demonstrated to possess a 5' exonuclease activity (Mathy, et al., 2007; Sharwood, et al., 2011). This observation together with the elucidation of the PPR10 mode of action (Pfalz, et al., 2009; Prikryl, et al., 2011) led A. Barkan (2011) to propose an alternative to the endoribonucleases-based model for intercistronic processing, i.e. site-specific blockage of both 5' and 3' exonucleases by a bound protein resulting in the formation of overlapping processed termini (Figure 4; Barkan, 2011). Yet, it has not been clarified if this model serves as a general mechanism for intercistronic processing. The high-resolution mapping of mRNA termini in this work revealed that the intercistronic processing in the barley *clpP-rps12 5', rps7-ndhB, rps4-ycf3* and *ndhE-psaC* regions should occur via a PPR10-like mechanism, since the 3' ends of the upstream RNAs were found to overlap with the 5' termini of the downstream ones and the shared sequences corresponded to a potential binding site of known or uncharacterized PPR/PPR-like protein (Figure 19, Appendix H). Similar observations were made for several other intercistronic mRNA termini in *Arabidopsis* (Ruwe and Schmitz-Linneweber, 2011). In contrast, no adjacent processed 5' and 3' termini mapping to intergenic regions (which would support the endoribonucleases-based model) were detected in both studies. Furthermore, a binding site of another PPR protein, i.e. HCF152, was shown to match the sequences shared by the 5' and 3' intercistronic termini stabilized by this protein (Zhelyazkova, et al., 2011). Taken together, these observations suggest that intercistronic processing in chloroplasts is mainly mediated *via* the site specific binding of a PPR/PPR-like protein to an intergenic region followed by 5' and/or 3' exonucleolytic digestion till the protein barrier.

4.3.3 Linking PPR footprints to RNA stability and translational enhancement

The binding site of PPR10 in *atpH* 5' UTR places it ~ 25 nt upstream of the *atpH* AUG. PPR10 was observed to stabilize processed *atpH* mRNAs as well as enhance *atpH* translation (Pfalz, et al., 2009). It was demonstrated that binding of a recombinant PPR10 to the *atpH* 5' UTR region disrupts the formation of a structure that sequesters a putative Shine-Dalgarno sequence. Thus, *atpH* translation enhancement was primarily attributed to the ability of a bound PPR10 to maintain the ribosome binding regions in a single stranded conformation. Furthermore, it was proposed that the placement of PPR10 in such a close proximity to the AUG could facilitate mRNA stability by minimizing the region in the 5'-UTR accessible to endonucleases (Prikryl, et al., 2011).) Our data indicates that the above described spatial relationship of a bound PPR to the start codon is a common feature (analysis performed by A.Barkan). The 3' boundaries of the majority (13 out of 19) of putative PPR/PPR-like binding sites mapped here in 5' UTRs are positioned ~ 20- to 60-nt upstream of the corresponding start codons. This positional bias allows PPR binding to occur without interfering with the access of ribosomes to the translation initiation region. Furthermore, it results in minimizing the length of 5' UTR regions that could be targeted by endonucleases. In addition, the putative PPR binding sites in *psbC* and *ndhK* 5' UTRs were predicted to form stable secondary structures including the Shine-Dalgarno sequences and thus result in translation inhibition (Figure 21; Zhelyazkova, et al., 2011). Therefore, the binding of a PPR protein to these regions would enhance translation by hindering the formation of these secondary structures. In addition, PPRs binding close to the start codon could in general reduce local RNA secondary structures and thus facilitate translation (Scharff, et al., 2011).

4.4 Concluding remarks

In this work, we have performed a comprehensive analysis of the primary and processed transcriptome of barley chloroplasts. The obtained results do not only contribute to the understanding of transcription regulation and RNA maturation, but will also serve as a basis for future work on the various aspects of gene expression regulation in plastids.

The genome-wide mapping of transcription start sites in mature chloroplasts and PEP-deficient plastids of the monocot plant barley revealed numerous general features of chloroplast transcription and shed light on polymerase specific gene expression in plastids. In order to learn

more about evolutionary conservation of the TSSs mapped in this study it will be necessary to investigate the transcriptome of dicot plastids using the same experimental and analytical approach in the future. In addition, such an analysis could give us new insights with respect to species-specific promoter usage and architecture. Moreover, the possibility to analyze NEP-mutants of dicots, in particular *Arabidopsis* lines with abolished RpoTp or RpoTmp, will further facilitate unraveling the division of labor between PEP and NEP in plastid transcription.

This study demonstrated the existence of numerous ncRNAs generated *via* transcription of free-standing genes in chloroplasts. However, the relevance and function of ncRNAs in plastids remains to be elucidated. In order to address this question, it will be worth determining the ncRNA composition in tobacco chloroplasts *in vivo* and to construct transplastomic plants lacking or over-expressing plastid ncRNAs based on the obtained information.

Last but not least, the results presented here provided evidence for protein-mediated protection from exonucleases being the predominant mechanism in processed mRNA formation. The list of potential PPR/PPR-like binding sites that was provided by this work is believed to facilitate the characterization of the corresponding proteins and the elucidation of their functions in chloroplast RNA processing.

5 REFERENCES

Allison, L. A.; Simon, L. D. and Maliga, P. (1996): Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants, EMBO (European Molecular Biology Organization) Journal 15 [11], pp. 2802-9.

Ayoubi, T. A. and Van De Ven, W. J. (1996): Regulation of gene expression by alternative promoters, FASEB Journal 10 [4], pp. 453-60.

Azevedo, J.; Courtois, F. and Lerbs-Mache, S. (2006): Sub-plastidial localization of two different phage-type RNA polymerases in spinach chloroplasts, Nucleic Acids Research 34 [2], pp. 436-44.

Baeza, L.; Bertrand, A.; Mache, R. and Lerbs-Mache, S. (1991): Characterization of a protein binding sequence in the promoter region of the 16S rRNA gene of the spinach chloroplast genome, Nucleic Acids Research 19 [13], pp. 3577-81.

Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J.; Li, W. W. and Noble, W. S. (2009): MEME SUITE: tools for motif discovery and searching, Nucleic Acids Research 37 [Web Server issue], pp. W202-8.

Ball, S.; Colleoni, C.; Cenci, U.; Raj, J. N. and Tirtiaux, C. (2010): The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis, J Exp Bot 62 [6], pp. 1775-801.

Barkan, A. (2011): Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold, Plant Physiology 155 [4], pp. 1520-32.

Barkan, A.; Walker, M.; Nolasco, M. and Johnson, D. (1994): A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms, EMBO (European Molecular Biology Organization) Journal 13 [13], pp. 3170-81.

Baumgartner, B. J.; Rapp, J. C. and Mullet, J. E. (1989): Plastid transcription activity and DNA copy number increase early in barley chloroplast development, Plant Physiology 89 [3], pp. 1011-8.

Baumgartner, B. J.; Rapp, J. C. and Mullet, J. E. (1993): Plastid Genes Encoding the Transcription/Translation Apparatus Are Differentially Transcribed Early in Barley (*Hordeum vulgare*) Chloroplast Development (Evidence for Selective Stabilization of *psbA* mRNA), Plant Physiology 101 [3], pp. 781-791.

Bendich, Arnold J. (1987): Why do chloroplasts and mitochondria contain so many copies of their genome?, BioEssays 6 [6], pp. 279-282.

Berends, T.; Gamble, P. E. and Mullet, J. E. (1987): Characterization of the barley chloroplast transcription units containing *psaA-psaB* and *psbD-psbC*, Nucleic Acids Research 15 [13], pp. 5217-40.

Berezikov, E.; Thuemmler, F.; van Laake, L. W.; Kondova, I.; Bontrop, R.; Cuppen, E. and Plasterk, R. H. (2006): Diversity of microRNAs in human and chimpanzee brain, Nature Genetics 38 [12], pp. 1375-7.

Bock, Ralph (2007): Structure, function, and inheritance of plastid genomes, Bock, Ralph, Cell and Molecular Biology of Plastids 19 pp. 29-63, Springer Berlin / Heidelberg.

Bollenbach, T. J.; Lange, H.; Gutierrez, R.; Erhardt, M.; Stern, D. B. and Gagliardi, D. (2005): RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*, Nucleic Acids Research 33 [8], pp. 2751-63.

Bollenbach, Thomas; Schuster, Gadi; Portnoy, Victoria and Stern, David (2007): Processing, degradation, and polyadenylation of chloroplast transcripts, Bock, Ralph, Cell and Molecular Biology of Plastids 19 pp. 175-211, Springer Berlin / Heidelberg.

Bompfunewerer, A. F.; Backofen, R.; Bernhart, S. H.; Hertel, J.; Hofacker, I. L.; Stadler, P. F. and Will, S. (2008): Variations on RNA folding and alignment: lessons from Benasque, J Math Biol 56 [1-2], pp. 129-44.

Boyer, S. K. and Mullet, J. E. (1988): Sequence and transcript map of barley chloroplast *psbA* gene, Nucleic Acids Research 16 [16], p. 8184.

Boyer, Scott K. and Mullet, John E. (1986): Characterization of *P. sativum* chloroplast *psbA* transcripts produced *in vivo*, *in vitro* and in *E. coli*, Plant Molecular Biology 6 [4], pp. 229-243.

Bradley, D. and Gatenby, A. A. (1985): Mutational analysis of the maize chloroplast ATPase-beta subunit gene promoter: the isolation of promoter mutants in *E. coli* and their characterization in a chloroplast *in vitro* transcription system, EMBO (European Molecular Biology Organization) Journal 4 [13B], pp. 3641-8.

Brantl, S. (2002): Antisense-RNA regulation and RNA interference, Biochim Biophys Acta 1575 [1-3], pp. 15-25.

Brantl, S. (2007): Regulatory mechanisms employed by *cis*-encoded antisense RNAs, Current Opinion in Microbiology 10 [2], pp. 102-9.

Briat, J. F.; Laulhere, J. P. and Mache, R. (1979): Transcription activity of a DNA-protein complex isolated from spinach plastids, European Journal of Biochemistry 98 [1], pp. 285-92.

Bünger, Wilhelm and Feierabend, J. (1980): Capacity for RNA synthesis in 70S ribosome-deficient plastids of heat-bleached rye leaves, Planta 149 [2], pp. 163-169.

Busch, A.; Richter, A. S. and Backofen, R. (2008): IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions, Bioinformatics 24 [24], pp. 2849-56.

Cahoon, A. B.; Harris, F. M. and Stern, D. B. (2004): Analysis of developing maize plastids reveals two mRNA stability classes correlating with RNA polymerase type, EMBO (European Molecular Biology Organization) Journal 5 [8], pp. 801-6.

Cai, W.; Okuda, K.; Peng, L. and Shikanai, T. (2011): PROTON GRADIENT REGULATION 3 recognizes multiple targets with limited similarity and mediates translation and RNA stabilization in plastids, The Plant Journal 67 [2], pp. 318-27.

Calie, P. J. and Manhart, J. R. (1994): Extensive sequence divergence in the 3' inverted repeat of the chloroplast *rbcL* gene in non-flowering land plants and algae, Gene 146 [2], pp. 251-6.

Canino, G.; Bocian, E.; Barbezier, N.; Echeverria, M.; Forner, J.; Binder, S. and Marchfelder, A. (2009): *Arabidopsis* encodes four tRNase Z enzymes, Plant Physiology 150 [3], pp. 1494-502.

Casano, L. M.; Martin, M. and Sabater, B. (2001): Hydrogen peroxide mediates the induction of chloroplastic Ndh complex under photooxidative stress in barley, Plant Physiology 125 [3], pp. 1450-8.

Cavalier-Smith, T. (2000): Membrane heredity and early chloroplast evolution, Trends in Plant Science 5 [4], pp. 174-82.

Chang, C. C.; Sheen, J.; Bligny, M.; Niwa, Y.; Lerbs-Mache, S. and Stern, D. B. (1999): Functional analysis of two maize cDNAs encoding T7-like RNA polymerases, Plant Cell 11 [5], pp. 911-26.

Chen, H. C. and Stern, D. B. (1991): Specific binding of chloroplast proteins in vitro to the 3' untranslated region of spinach chloroplast *petD* mRNA, Molecular and Cellular Biology 11 [9], pp. 4380-8.

Cheng, Y. S.; Lin, C. H. and Chen, L. J. (1997): Transcription and processing of the gene for spinach chloroplast threonine tRNA in a homologous *in vitro* system, Biochemical and Biophysical Research Communications 233 [2], pp. 380-5.

Christopher, D. A.; Kim, M. and Mullet, J. E. (1992): A novel light-regulated promoter is conserved in cereal and dicot chloroplasts, Plant Cell 4 [7], pp. 785-98.

Courtois, F.; Merendino, L.; Demarsy, E.; Mache, R. and Lerbs-Mache, S. (2007): Phage-type RNA polymerase RPOTmp transcribes the *rrn* operon from the PC promoter at early developmental stages in *Arabidopsis*, Plant Physiology 145 [3], pp. 712-21.

Crooks, G. E.; Hon, G.; Chandonia, J. M. and Brenner, S. E. (2004): WebLogo: a sequence logo generator, Genome Research 14 [6], pp. 1188-90.

Deana, A.; Celesnik, H. and Belasco, J. G. (2008): The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal, Nature 451 [7176], pp. 355-8.

Dekker, J. P. and Boekema, E. J. (2005): Supramolecular organization of thylakoid membrane proteins in green plants, Biochim Biophys Acta 1706 [1-2], pp. 12-39.

Del Campo, E. M.; Sabater, B. and Martin, M. (2002): Post-transcriptional control of chloroplast gene expression. Accumulation of stable *psaC* mRNA is due to downstream RNA cleavages in the *ndhD* gene, Journal of Biological Chemistry 277 [39], pp. 36457-64.

Demarsy, E.; Buhr, F.; Lambert, E. and Lerbs-Mache, S. (2011): Characterization of the plastid-specific germination and seedling establishment transcriptional programme, J Exp Bot 63 [2], pp. 925-39.

Demarsy, E.; Courtois, F.; Azevedo, J.; Buhot, L. and Lerbs-Mache, S. (2006): Building up of the plastid transcriptional machinery during germination and early plant development, Plant Physiology 142 [3], pp. 993-1003.

Duhring, U.; Axmann, I. M.; Hess, W. R. and Wilde, A. (2006): An internal antisense RNA regulates expression of the photosynthesis gene *isiA*, Proceedings of the National Academy of Sciences of the United States of America 103 [18], pp. 7054-8.

Dyall, S. D.; Brown, M. T. and Johnson, P. J. (2004): Ancient invasions: from endosymbionts to organelles, Science 304 [5668], pp. 253-7.

Eisermann, A.; Tiller, K. and Link, G. (1990): In vitro transcription and DNA binding characteristics of chloroplast and etioplast extracts from mustard (Sinapis alba) indicate differential usage of the *psbA* promoter, EMBO (European Molecular Biology Organization) Journal 9 [12], pp. 3981-7.

Ellis, R. J. and Hartley, M. R. (1971): Sites of synthesis of chloroplast proteins, Nature 233 [5320], pp. 193-6.

Emanuel, C.; von Groll, U.; Müller, M.; Börner, T. and Weihe, A. (2006): Development- and tissue-specific expression of the RpoT gene family of *Arabidopsis* encoding mitochondrial and plastid RNA polymerases, Planta 223 [5], pp. 998-1009.

Emanuel, C.; Weihe, A.; Graner, A.; Hess, W. R. and Börner, T. (2004): Chloroplast development affects expression of phage-type RNA polymerases in barley leaves, The Plant Journal 38 [3], pp. 460-72.

Ems, S. C.; Morden, C. W.; Dixon, C. K.; Wolfe, K. H.; dePamphilis, C. W. and Palmer, J. D. (1995): Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*, Plant Molecular Biology 29 [4], pp. 721-33.

Felder, S.; Meierhoff, K.; Sane, A. P.; Meurer, J.; Driemel, C.; Plucken, H.; Klaff, P.; Stein, B.; Bechtold, N. and Westhoff, P. (2001): The nucleus-encoded *HCF107* gene of *Arabidopsis* provides a link between intercistronic RNA processing and the accumulation of translation-competent *psbH* transcripts in chloroplasts, Plant Cell 13 [9], pp. 2127-41.

Fire, A. (1999): RNA-triggered gene silencing, Trends in Genetics 15 [9], pp. 358-63.

Fisk, D. G.; Walker, M. B. and Barkan, A. (1999): Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression, EMBO (European Molecular Biology Organization) Journal 18 [9], pp. 2621-30.

Franco-Zorrilla, J. M.; Valli, A.; Todesco, M.; Mateos, I.; Puga, M. I.; Rubio-Somoza, I.; Leyva, A.; Weigel, D.; Garcia, J. A. and Paz-Ares, J. (2007): Target mimicry provides a new mechanism for regulation of microRNA activity, Nature Genetics 39 [8], pp. 1033-7.

Gatenby, A. A.; Castleton, J. A. and Saul, M. W. (1981): Expression in *E. coli* of maize and wheat chloroplast genes for large subunit of ribulose bisphosphate carboxylase, Nature 291 [5811], pp. 117-21.

Georg, J. and Hess, W. R. (2011): Regulatory RNAs in cyanobacteria: developmental decisions, stress responses and a plethora of chromosomally encoded cis-antisense RNAs, Biological Chemistry 392 [4], pp. 291-7.

Georg, J.; Honsel, A.; Voss, B.; Rennenberg, H. and Hess, W. R. (2010): A long antisense RNA in plant chloroplasts, New Phytologist 186 [3], pp. 615-22.

Georg, J.; Voss, B.; Scholz, I.; Mitschke, J.; Wilde, A. and Hess, W. R. (2009): Evidence for a major role of antisense RNAs in cyanobacterial gene regulation, Molecular System Biology 5, p. 305.

Germain, A.; Herlich, S.; Larom, S.; Kim, S. H.; Schuster, G. and Stern, D. B. (2011): Mutational analysis of *Arabidopsis* chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation, The Plant Journal 67 [3], pp. 381-94.

Goldman, Seth R; Sharp, Josh S; Vvedenskaya, Irina O; Livny, Jonathan; Dove, Simon L and Nickels, Bryce E (2011): NanoRNAs Prime Transcription Initiation In Vivo, Molecular Cell 42 [6], pp. 817-825.

Gould, S. B.; Waller, R. F. and McFadden, G. I. (2008): Plastid evolution, Annual Review of Plant Biology 59, pp. 491-517.

Green, B. R. (2011): Chloroplast genomes of photosynthetic eukaryotes, The Plant Journal 66 [1], pp. 34-44.

Gruissem, W.; Elsner-Menzel, C.; Latshaw, S.; Narita, J. O.; Schaffer, M. A. and Zurawski, G. (1986): A subpopulation of spinach chloroplast tRNA genes does not require upstream promoter elements for transcription, Nucleic Acids Research 14 [19], pp. 7541-56.

Gruissem, W. and Zurawski, G. (1985): Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene, EMBO (European Molecular Biology Organization) Journal 4 [7], pp. 1637-44.

Gruissem, Wilhelm; Barkan, Alice; Deng, Xing-wang and Stern, David (1988): Transcriptional and post-transcriptional control of plastid mRNA levels in higher plants, Trends in Genetics 4 [9], pp. 258-263.

Guertin, M. and Bellemare, G. (1979): Synthesis of chloroplast ribonucleic acid in *Chlamydomonas reinhardtii* toluene-treated cells, European Journal of Biochemistry 96 [1], pp. 125-9.

Hajdukiewicz, P. T.; Allison, L. A. and Maliga, P. (1997): The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids, EMBO (European Molecular Biology Organization) Journal 16 [13], pp. 4041-8.

Han, Chang-deok; Patrie, William; Polacco, Mary and Coe, Edward H. (1993): Aberrations in plastid transcripts and deficiency of plastid DNA in striped and albino mutants in maize, Planta 191 [4], pp. 552-563.

Hanaoka, M.; Kanamaru, K.; Takahashi, H. and Tanaka, K. (2003): Molecular genetic analysis of chloroplast gene promoters dependent on SIG2, a nucleus-encoded sigma factor for the plastid-encoded RNA polymerase, in *Arabidopsis thaliana*, Nucleic Acids Research 31 [24], pp. 7090-8.

Hao, Weilong and Palmer, Jeffrey D. (2009): Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes, Proceedings of the National Academy of Sciences of the United States of America 106 [39], pp. 16728-16733.

Harley, C. B. and Reynolds, R. P. (1987): Analysis of *E. coli* promoter sequences, Nucleic Acids Research 15 [5], pp. 2343-61.

Hashimoto, M.; Endo, T.; Peltier, G.; Tasaka, M. and Shikanai, T. (2003): A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in *Arabidopsis*, The Plant Journal 36 [4], pp. 541-9.

Hattori, M.; Miyake, H. and Sugita, M. (2007): A Pentatricopeptide repeat protein is required for RNA processing of clpP Pre-mRNA in moss chloroplasts, Journal of Biological Chemistry 282 [14], pp. 10773-82.

Hattori, M. and Sugita, M. (2009): A moss pentatricopeptide repeat protein binds to the 3' end of plastid *clpP* pre-mRNA and assists with mRNA maturation, FEBS Journal 276 [20], pp. 5860-9.

He, L. and Hannon, G. J. (2004): MicroRNAs: small RNAs with a big role in gene regulation, Nature Reviews Genetics 5 [7], pp. 522-31.

Hedtke, B.; Borner, T. and Weihe, A. (1997): Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*, Science 277 [5327], pp. 809-11.

Hedtke, B.; Borner, T. and Weihe, A. (2000): One RNA polymerase serving two genomes, EMBO Reports 1 [5], pp. 435-40.

Hedtke, B.; Meixner, M.; Gillandt, S.; Richter, E.; Borner, T. and Weihe, A. (1999): Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo, The Plant Journal 17 [5], pp. 557-61.

Hegeman, C. E.; Halter, C. P.; Owens, T. G. and Hanson, M. R. (2005): Expression of complementary RNA from chloroplast transgenes affects editing efficiency of transgene and endogenous chloroplast transcripts, Nucleic Acids Research 33 [5], pp. 1454-64.

Hess, W. R. and Börner, T. (1999): Organellar RNA polymerases of higher plants, Int Rev Cytol 190, pp. 1-59.

Hess, W. R.; Hoch, B.; Zeltz, P.; Hübschmann, T.; Kossel, H. and Börner, T. (1994): Inefficient *rpl2* splicing in barley mutants with ribosome-deficient plastids, Plant Cell 6 [10], pp. 1455-65.

Hess, W. R.; Prombona, A.; Fieder, B.; Subramanian, A. R. and Börner, T. (1993): Chloroplast *rps15* and the *rpoB*/*C1*/*C2* gene cluster are strongly transcribed in ribosome-deficient plastids: evidence for a functioning non-chloroplast-encoded RNA polymerase, EMBO (European Molecular Biology Organization) Journal 12 [2], pp. 563-71.

Hildebrand, M.; Hallick, R. B.; Passavant, C. W. and Bourque, D. P. (1988): Trans-splicing in chloroplasts: the *rps12* loci of *Nicotiana tabacum*, Proceedings of the National Academy of Sciences of the United States of America 85 [2], pp. 372-6.

Hirose, Tetsuro and Sugiura, Masahiro (1997): Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast *ndhD* mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes, EMBO (European Molecular Biology Organization) Journal 16 [22], pp. 6804-6811.

Hofacker, I. L.; Fontana, W.; Stadler, P. F.; Bonhoeffer, L. S.; Tacker, M. and Schuster, P. (1994): Fast folding and comparison of RNA secondary structures, Monatshefte f. Chemie 125 [2], pp. 167-188.

Hofacker, I. L. and Stadler, P. F. (2006): Memory efficient folding algorithms for circular RNA secondary structures, Bioinformatics 22 [10], pp. 1172-6.

Hotto, A. M.; Huston, Z. E. and Stern, D. B. (2010): Overexpression of a natural chloroplast-encoded antisense RNA in tobacco destabilizes 5S rRNA and retards plant growth, BMC Plant Biology 10, p. 213.

Hotto, Amber M.; Schmitz, Robert J.; Fei, Zhangjun; Ecker, Joseph R. and Stern, David B. (2011): Unexpected Diversity of Chloroplast Noncoding RNAs as Revealed by Deep Sequencing of the *Arabidopsis* Transcriptome, G3: Genes, Genomes, Genetics 1 [7], pp. 559-570.

Hricova, A.; Quesada, V. and Micol, J. L. (2006): The SCABRA3 nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in *Arabidopsis*, Plant Physiology 141 [3], pp. 942-56.

Hübschmann, T. and Börner, T. (1998): Characterisation of transcript initiation sites in ribosome-deficient barley plastids, Plant Molecular Biology 36 [3], pp. 493-6.

Hübschmann, T.; Hess, W. R. and Börner, T. (1996): Impaired splicing of the *rps12* transcript in ribosome-deficient plastids, Plant Molecular Biology 30 [1], pp. 109-23.

Hudson, G. S.; Holton, T. A.; Whitfield, P. R. and Bottomley, W. (1988): Spinach chloroplast *rpoBC* genes encode three subunits of the chloroplast RNA polymerase, Journal of Molecular Biology 200 [4], pp. 639-54.

Huttenhofer, A. and Noller, H. F. (1994): Footprinting mRNA-ribosome complexes with chemical probes, EMBO (European Molecular Biology Organization) Journal 13 [16], pp. 3892-901.

Inada, H.; Seki, M.; Morikawa, H.; Nishimura, M. and Iba, K. (1997): Existence of three regulatory regions each containing a highly conserved motif in the promoter of plastid-encoded RNA polymerase gene (*rpoB*), The Plant Journal 11 [4], pp. 883-90.

Iratni, R.; Baeza, L.; Andreeva, A.; Mache, R. and Lerbs-Mache, S. (1994): Regulation of rDNA transcription in chloroplasts: promoter exclusion by constitutive repression, Genes Dev 8 [23], pp. 2928-38.

Iratni, R.; Diederich, L.; Harrak, H.; Bligny, M. and Lerbs-Mache, S. (1997): Organ-specific transcription of the *rrn* operon in spinach plastids, Journal of Biological Chemistry 272 [21], pp. 13676-82.

Isono, K.; Niwa, Y.; Satoh, K. and Kobayashi, H. (1997): Evidence for transcriptional regulation of plastid photosynthesis genes in *Arabidopsis thaliana* roots, Plant Physiology 114 [2], pp. 623-30.

Jacob, FranÃ§ois and Monod, Jacques (1961): On the Regulation of Gene Activity, Cold Spring Harbor Symposia on Quantitative Biology 26, pp. 193-211.

Jäger, D.; Sharma, C. M.; Thomsen, J.; Ehlers, C.; Vogel, J. and Schmitz, R. A. (2009): Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability, Proceedings of the National Academy of Sciences of the United States of America 106 [51], pp. 21878-82.

Jahn, Dieter (1992): Expression of the *Chlamydomonas reinhardtii* chloroplast $tRNA^{Glu}$ gene in a homologous *in vitro* transcription system is independent of upstream promoter elements Archives of Biochemistry and Biophysics 298 [2], pp. 505-513.

Johnson, C.; Bowman, L.; Adai, A. T.; Vance, V. and Sundaresan, V. (2007): CSRDB: a small RNA integrated database and browser resource for cereals, Nucleic Acids Research 35 [Database issue], pp. D829-33.

Johnson, X.; Wostrikoff, K.; Finazzi, G.; Kuras, R.; Schwarz, C.; Bujaldon, S.; Nickelsen, J.; Stern, D. B.; Wollman, F. A. and Vallon, O. (2010): MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in *Chlamydomonas* and *Arabidopsis*, Plant Cell 22 [1], pp. 234-48.

Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosawa, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M. and Tabata, S. (1996): Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement), DNA Res 3 [3], pp. 185-209.

Kanervo, Eira; Suorsa, Marjaana and Aro, Eva-Mari (2007): Assembly of protein complexes in plastids, Bock, Ralph, Cell and Molecular Biology of Plastids 19 pp. 283-313, Springer Berlin / Heidelberg.

Kanno, Akira and Hirai, Atsushi (1993): A transcription map of the chloroplast genome from rice (*Oryza sativa*), Current Genetics 23 [2], pp. 166-174.

Kapoor, S. and Sugiura, M. (1999): Identification of two essential sequence elements in the nonconsensus type II P*atpB*-290 plastid promoter by using plastid transcription extracts from cultured tobacco BY-2 cells, Plant Cell 11 [9], pp. 1799-810.

Kapoor, S.; Suzuki, J. Y. and Sugiura, M. (1997): Identification and functional significance of a new class of non-consensus-type plastid promoters, The Plant Journal 11 [2], pp. 327-37.

Kawaguchi, H.; Fukuda, I.; Shiina, T. and Toyoshima, Y. (1992): Dynamical behavior of *psb* gene transcripts in greening wheat seedlings. I. Time course of accumulation of the *psbA* through *psbN* gene transcripts during light-induced greening, Plant Molecular Biology 20 [4], pp. 695-704.

Kim, J. W.; Park, J. K.; Kim, B. H.; Lee, J. S. and Sim, W. S. (2002): Molecular analysis of the accumulation of the transcripts of the large subunit gene of ribulose-1,5-bisphosphate carboxylase/oxygenase by light, Mol Cells 14 [2], pp. 281-7.

Kim, M. and Mullet, J. E. (1995): Identification of a sequence-specific DNA binding factor required for transcription of the barley chloroplast blue light-responsive *psbD-psbC* promoter, Plant Cell 7 [9], pp. 1445-57.

Kim, M.; Thum, K. E.; Morishige, D. T. and Mullet, J. E. (1999): Detailed architecture of the barley chloroplast *psbD-psbC* blue light-responsive promoter, Journal of Biological Chemistry 274 [8], pp. 4684-92.

Kim, Minkyun; Christopher, David and Mullet, John (1993): Direct evidence for selective modulation of *psbA*, *rpoA*, *rbcL* and 16S RNA stability during barley chloroplast development, Plant Molecular Biology 22 [3], pp. 447-463.

Kleine, T.; Maier, U. G. and Leister, D. (2009): DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis, Annual Review of Plant Biology 60, pp. 115-38.

Kobayashi, Y.; Dokiya, Y. and Sugita, M. (2001): Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts, Biochem Biophys Res Commun 289 [5], pp. 1106-13.

Krause, K.; Maier, R. M.; Kofer, W.; Krupinska, K. and Herrmann, R. G. (2000): Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome, Molecular and General Genetics 263 [6], pp. 1022-30.

Krause, Kirsten and Krupinska, Karin (2000): Molecular and functional properties of highly purified transcriptionally active chromosomes from spinach chloroplasts, Physiologia Plantarum 109 [2], pp. 188-195.

Kühn, K.; Bohne, A. V.; Liere, K.; Weihe, A. and Börner, T. (2007): Arabidopsis phage-type RNA polymerases: accurate in vitro transcription of organellar genes, Plant Cell 19 [3], pp. 959-71.

Kühn, K.; Weihe, A. and Börner, T. (2005): Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*, Nucleic Acids Research 33 [1], pp. 337-46.

Kuroda, H.; Suzuki, H.; Kusumegi, T.; Hirose, T.; Yukawa, Y. and Sugiura, M. (2007): Translation of *psbC* mRNAs starts from the downstream GUG, not the upstream AUG, and requires the extended Shine-Dalgarno sequence in tobacco chloroplasts, Plant Cell Physiology 48 [9], pp. 1374-8.

Kuroiwa, T. (1991): The replication, differentiation, and inheritance of plastids with emphasis on the concept of organelle nuclei, International Review of Cytology-a Survey of Cell Biology 128, pp. 1-

Legen, J.; Kemp, S.; Krause, K.; Profanter, B.; Herrmann, R. G. and Maier, R. M. (2002): Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries, The Plant Journal 31 [2], pp. 171-88.

Lerbs-Mache, S. (2011): Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription?, Plant Molecular Biology 76 [3-5], pp. 235-49.

Liere, K.; Kaden, D.; Maliga, P. and Borner, T. (2004): Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type, Nucleic Acids Research 32 [3], pp. 1159-65.

Liere, K. and Link, G. (1994): Structure and expression characteristics of the chloroplast DNA region containing the split gene for $tRNA^{(Gly)}$ (UCC) from mustard (*Sinapis alba* L.), Current Genetics 26 [5-6], pp. 557-63.

Liere, K. and Link, G. (1995): RNA-binding activity of the *matK* protein encoded by the chloroplast trnK intron from mustard (*Sinapis alba* L.), Nucleic Acids Research 23 [6], pp. 917-21.

Liere, Karsten and Börner, Thomas (2007): Transcription and transcriptional regulation in plastids, Bock, Ralph, Cell and Molecular Biology of Plastids 19 pp. 121-174, Springer Berlin / Heidelberg.

Liere, Karsten and Maliga, Pal (1999): In vitro characterization of the tobacco *rpoB* promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters, EMBO (European Molecular Biology Organization) Journal 18 [1], pp. 249-257.

Liere, Karsten and Maliga, Pal (2001): Plastid RNA Polymerases in Higher Plants, Aro, Eva-Mari and Andersson, Bertil, Regulation of Photosynthesis 11 pp. 29-49, Springer Netherlands.

Liere, Karsten; Weihe, Andreas and Börner, Thomas (2011): The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation, Journal of Plant Physiology 168 [12], pp. 1345-1360.

Lilly, J. W.; Havey, M. J.; Jackson, S. A. and Jiang, J. (2001): Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants, Plant Cell 13 [2], pp. 245-54.

Link, G. (1984): DNA sequence requirements for the accurate transcription of a protein-coding plastid gene in a plastid in vitro system from mustard (*Sinapis alba L.*), EMBO (European Molecular Biology Organization) Journal 3 [8], pp. 1697-704.

Lioliou, E.; Romilly, C.; Romby, P. and Fechter, P. (2010): RNA-mediated regulation in bacteria: from natural to artificial systems, New Biotechnology 27 [3], pp. 222-35.

Liu, B. and Troxler, R. F. (1996): Molecular characterization of a positively photoregulated nuclear gene for a chloroplast RNA polymerase sigma factor in *Cyanidium caldarium*, Proceedings of the National Academy of Sciences of the United States of America 93 [8], pp. 3313-8.

Liu, J. W. and Rose, R. J. (1992): The spinach chloroplast chromosome is bound to the thylakoid membrane in the region of the inverted repeat, Biochem Biophys Res Commun 184 [2], pp. 993-1000.

Lohse, Marc; Drechsel, Oliver and Bock, Ralph (2007): OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes, Current Genetics 52 [5], pp. 267-274.

Loiselay, C.; Gumpel, N. J.; Girard-Bascou, J.; Watson, A. T.; Purton, S.; Wollman, F. A. and Choquet, Y. (2008): Molecular identification and function of *cis*- and *trans*-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts, Molecular and Cellular Biology 28 [17], pp. 5529-42.

Lopez-Juez, E. and Pyke, K. A. (2005): Plastids unleashed: their development and their integration in plant development, International Journal of Developmental Biology 49 [5-6], pp. 557-77.

Lowe, T. M. and Eddy, S. R. (1997): tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Research 25 [5], pp. 955-64.

Lung, B.; Zemann, A.; Madej, M. J.; Schuelke, M.; Techritz, S.; Ruf, S.; Bock, R. and Huttenhofer, A. (2006): Identification of small non-coding RNAs from mitochondria and chloroplasts, Nucleic Acids Research 34 [14], pp. 3842-52.

Lyubetsky, V. A.; Zverkov, O. A.; Rubanov, L. I. and Seliverstov, A. V. (2011): Modeling RNA polymerase competition: the effect of sigma-subunit knockout and heat shock on gene transcription level, Biology Direct 6, p. 3.

Maier, U. G.; Bozarth, A.; Funk, H. T.; Zauner, S.; Rensing, S. A.; Schmitz-Linneweber, C.; Borner, T. and Tillich, M. (2008): Complex chloroplast RNA metabolism: just debugging the genetic programme?, BMC Biol 6, p. 36.

María del Campo, Eva; Sabater, Bartolomé and Martín, Mercedes (2006): Characterization of the 5'- and 3'-ends of mRNAs of *ndhH*, *ndhA* and *ndhI* genes of the plastid *ndhH-D* operon, Biochimie 88 [3-4], pp. 347-357.

Marín-Navarro, Julia; Manuell, Andrea; Wu, Joann and P. Mayfield, Stephen (2007): Chloroplast translation regulation, Photosynthesis Research 94 [2], pp. 359-374.

Martin, W. and Herrmann, R. G. (1998): Gene transfer from organelles to the nucleus: how much, what happens, and why?, Plant Physiology 118 [1], pp. 9-17.

Mathews, D. E. and Durbin, R. D. (1990): Tagetitoxin inhibits RNA synthesis directed by RNA polymerases from chloroplasts and *Escherichia coli*, Journal of Biological Chemistry 265 [1], pp. 493-8.

Mathy, N.; Benard, L.; Pellegrini, O.; Daou, R.; Wen, T. and Condon, C. (2007): 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA, Cell 129 [4], pp. 681-92.

Mattoo, A. K.; Marder, J. B. and Edelman, M. (1989): Dynamics of the photosystem II reaction center, Cell 56 [2], pp. 241-6.

McCaskill, J. S. (1990): The equilibrium partition function and base pair binding probabilities for RNA secondary structure, Biopolymers 29 [6-7], pp. 1105-19.

Meierhoff, K.; Felder, S.; Nakamura, T.; Bechtold, N. and Schuster, G. (2003): HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs, Plant Cell 15 [6], pp. 1480-95.

Memon, Abdul R.; Meng, Bingyuan and Mullet, John E. (1996): RNA-binding proteins of 37/38 kDa bind specifically to the barley chloroplast *psbA* 3′-end untranslated RNA, Plant Molecular Biology 30 [6], pp. 1195-1205.

Mereschkowsky, C (1910): Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen, Biologisches Centralblatt 30, pp. 277-303.

Michel, F. and Ferat, J. L. (1995): Structure and activities of group II introns, Annu Rev Biochem 64, pp. 435-61.

Mitschke, J.; Georg, J.; Scholz, I.; Sharma, C. M.; Dienst, D.; Bantscheff, J.; Voss, B.; Steglich, C.; Wilde, A.; Vogel, J. and Hess, W. R. (2011): An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803, Proceedings of the National Academy of Sciences of the United States of America 108 [5], pp. 2124-9.

Mohorianu, I.; Schwach, F.; Jing, R.; Lopez-Gomollon, S.; Moxon, S.; Szittya, G.; Sorefan, K.; Moulton, V. and Dalmay, T. (2011): Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns, The Plant Journal 67 [2], pp. 232-46.

Morden, C. W.; Delwiche, C. F.; Kuhsel, M. and Palmer, J. D. (1992): Gene phylogenies and the endosymbiotic origin of plastids, Biosystems 28 [1-3], pp. 75-90.

Morin, R. D.; Aksay, G.; Dolgosheina, E.; Ebhardt, H. A.; Magrini, V.; Mardis, E. R.; Sahinalp, S. C. and Unrau, P. J. (2008): Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*, Genome Research 18 [4], pp. 571-84.

Mudd, E. A.; Sullivan, S.; Gisby, M. F.; Mironov, A.; Kwon, C. S.; Chung, W. I. and Day, A. (2008): A 125 kDa RNase E/G-like protein is present in plastids and is essential for chloroplast development and autotrophic growth in *Arabidopsis*, J Exp Bot 59 [10], pp. 2597-610.

Mullet, J. E. (1993): Dynamic regulation of chloroplast transcription, Plant Physiology 103 [2], pp. 309-13.

Nass, M. M. and Nass, S. (1963): Intramitochondrial Fibers with DNA Characteristics. I. Fixation and Electron Staining Reactions, J Cell Biol 19, pp. 593-611.

Neuhaus, H. E. and Emes, M. J. (2000): Nonphotosynthetic Metabolism in Plastids, Annual Review of Plant Physiology and Plant Molecular Biology 51, pp. 111-140.

Neuhaus, H. and Link, G. (1990): The chloroplast *psbK* operon from mustard (*Sinapis alba* L.): multiple transcripts during seedling development and evidence for divergent overlapping transcription, Current Genetics 18 [4], pp. 377-83.

Nickelsen, J. and Link, G. (1990): Nucleotide sequence of the mustard chloroplast genes *trnH* and *rps19*, Nucleic Acids Research 18 [4], p. 1051.

Nisbet, E. G. and Sleep, N. H. (2001): The habitat and nature of early life, Nature 409 [6823], pp. 1083-1091.

Nishimura, Y.; Kikis, E. A.; Zimmer, S. L.; Komine, Y. and Stern, D. B. (2004): Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in chlamydomonas chloroplasts, Plant Cell 16 [11], pp. 2849-69.

Noutsos, C.; Richly, E. and Leister, D. (2005): Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants, Genome Research 15 [5], pp. 616-28.

O'Toole, N.; Hattori, M.; Andres, C.; Iida, K.; Lurin, C.; Schmitz-Linneweber, C.; Sugita, M. and Small, I. (2008): On the expansion of the pentatricopeptide repeat gene family in plants, Mol Biol Evol 25 [6], pp. 1120-8.

Ogawa, T.; Yoshimura, K.; Miyake, H.; Ishikawa, K.; Ito, D.; Tanabe, N. and Shigeoka, S. (2008): Molecular characterization of organelle-type Nudix hydrolases in *Arabidopsis*, Plant Physiology 148 [3], pp. 1412-24.

Ogrzewalla, K.; Piotrowski, M.; Reinbothe, S. and Link, G. (2002): The plastid transcription kinase from mustard (*Sinapis alba L.*). A nuclear-encoded CK2-type chloroplast enzyme with redox-sensitive function, European Journal of Biochemistry 269 [13], pp. 3329-37.

Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Sano, T.; Sano, S.; Shirai, H.; Umesono, K.; Shiki, Y.; Takeuchi, M.; Chang, Z. and et al. (1988): Structure and organization of *Marchantia polymorpha* chloroplast genome. I. Cloning and gene identification, Journal of Molecular Biology 203 [2], pp. 281-98.

Ohyama, Kanji; Fukuzawa, Hideya; Kohchi, Takayuki; Shirai, Hiromasa; Sano, Tohru; Sano, Satoshi; Umesono, Kazuhiko; Shiki, Yasuhiko; Takeuchi, Masayuki; Chang, Zhen; Aota, Shin-ichi; Inokuchi, Hachiro and Ozeki, Haruo (1986): Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA, Nature 322 [6079], pp. 572-574.

Oldenburg, D. J. and Bendich, A. J. (2004): Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms, Journal of Molecular Biology 335 [4], pp. 953-70.

Oliver, Richard and Poulsen, Carsten (1984): Structure of a heavily transcribed region of barley chloroplast DNA. Transfer RNA genes for serine (UGA), glycine (GCC, UCC), formyl-methionine and threonine (GGU), Carlsberg research communications 49 [7], pp. 647-673.

Opdyke, J. A.; Kang, J. G. and Storz, G. (2004): GadY, a small-RNA regulator of acid response genes in *Escherichia coli*, J Bacteriol 186 [20], pp. 6698-705.

Ozsolak, F. and Milos, P. M. (2011): RNA sequencing: advances, challenges and opportunities, Nature Reviews Genetics 12 [2], pp. 87-98.

Papenfort, K. and Vogel, J. (2009): Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level, Research in Microbiology 160 [4], pp. 278-87.

Passalacqua, K. D.; Varadarajan, A.; Ondov, B. D.; Okou, D. T.; Zwick, M. E. and Bergman, N. H. (2009): Structure and complexity of a bacterial transcriptome, J Bacteriol 191 [10], pp. 3203-11.

Peled-Zehavi, Hadas and Danon, Avihai (2007): Translation and translational regulation in chloroplasts, Bock, Ralph 19 pp. 249-281, Springer Berlin / Heidelberg.

Pfalz, J.; Bayraktar, O. A.; Prikryl, J. and Barkan, A. (2009): Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts, EMBO (European Molecular Biology Organization) Journal 28 [14], pp. 2042-52.

Pfalz, J.; Liere, K.; Kandlbinder, A.; Dietz, K. J. and Oelmuller, R. (2006): pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression, Plant Cell 18 [1], pp. 176-97.

Pfannschmidt, T. and Link, G. (1997): The A and B forms of plastid DNA-dependent RNA polymerase from mustard (*Sinapis alba L.*) transcribe the same genes in a different developmental context, Molecular and General Genetics 257 [1], pp. 35-44.

Pfannschmidt, T.; Ogrzewalla, K.; Baginsky, S.; Sickmann, A.; Meyer, H. E. and Link, G. (2000): The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba L.*). Integration of a prokaryotic core into a larger complex with organelle-specific functions, European Journal of Biochemistry 267 [1], pp. 253-61.

Pinto, A. C.; Melo-Barbosa, H. P.; Miyoshi, A.; Silva, A. and Azevedo, V. (2011): Application of RNA-seq to reveal the transcript profile in bacteria, Genet Mol Res 10 [3], pp. 1707-18.

Poulsen, Carsten (1984): Two mRNA species differing by 258 nucleotides at the 5′ end are formed from the barley chloroplast *rbcL* gene, Carlsberg Research Communications 49 [1], pp. 89-104.

Prasanth, K. V. and Spector, D. L. (2007): Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum, Genes Dev 21 [1], pp. 11-42.

Prikryl, J.; Rojas, M.; Schuster, G. and Barkan, A. (2011): Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein, Proceedings of the National Academy of Sciences of the United States of America 108 [1], pp. 415-20.

Purton, S. and Gray, J. C. (1989): The plastid *rpoA* gene encoding a protein homologous to the bacterial RNA polymerase alpha subunit is expressed in pea chloroplasts, Molecular and General Genetics 217 [1], pp. 77-84.

Pyke, Kevin (2007): Plastid biogenesis and differentiation, Bock, Ralph, Cell and Molecular Biology of Plastids 19 pp. 1-28, Springer Berlin / Heidelberg.

Reinbothe, S.; Reinbothe, C.; Heintzen, C.; Seidenbecher, C. and Parthier, B. (1993): A methyl jasmonate-induced shift in the length of the 5' untranslated region impairs translation of the plastid *rbcL* transcript in barley, EMBO (European Molecular Biology Organization) Journal 12 [4], pp. 1505-12.

Repoila, F. and Darfeuille, F. (2009): Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects, Biol Cell 101 [2], pp. 117-31.

Richards, Jamie; Liu, Quansheng; Pellegrini, Olivier; Celesnik, Helena; Yao, Shiyi; Bechhofer, David H.; Condon, Ciaran and Belasco, Joel G. (2011): An RNA Pyrophosphohydrolase Triggers 5'-Exonucleolytic Degradation of mRNA in *Bacillus subtilis*, Molecular Cell 43 [6], pp. 940-949.

Richter, A. S.; Schleberger, C.; Backofen, R. and Steglich, C. (2009): Seed-based INTARNA prediction combined with GFP-reporter system identifies mRNA targets of the small RNA Yfr1, Bioinformatics 26 [1], pp. 1-5.

Ris, H. and Plaut, W. (1962): Ultrastructure of DNA-containing areas in the chloroplast of *Chlamydomonas*, J Cell Biol 13, pp. 383-91.

Rozen, S. and Skaletsky, H. (2000): Primer3 on the WWW for general users and for biologist programmers, Methods in Molecular Biology 132, pp. 365-86.

Ruwe, H. and Schmitz-Linneweber, C. (2011): Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins?, Nucleic Acids Research.

Sagan, Lynn (1967): On the origin of mitosing cells, Journal of Theoretical Biology 14 [3], pp. 225-IN6.

Saha, D.; Prasad, A. M. and Srinivasan, R. (2007): Pentatricopeptide repeat proteins and their emerging roles in plants, Plant Physiology and Biochemistry 45 [8], pp. 521-34.

Sakai, Atsushi; Yamashita, Hirofumi; Nemoto, Yasuyuki; Kawano, Shigeyuki and Kuroiwa, Tsuneyoshi (1991): Transcriptional Activity of Morphologically Intact Proplastid-Nuclei (Nucleoids) Isolated from Tobacco Cultured Cells, Plant and Cell Physiology 32 [6], pp. 835-843.

Sambrook, Joseph and Russell, David W. (2001): Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., ISBN: 0879695773 (pbk.)
0879695765 (hbk.).

Sane, A. P.; Stein, B. and Westhoff, P. (2005): The nuclear gene HCF107 encodes a membrane-associated R-TPR (RNA tetratricopeptide repeat)-containing protein involved in expression of the plastidial *psbH* gene in *Arabidopsis*, The Plant Journal 42 [5], pp. 720-30.

Saski, C.; Lee, S. B.; Fjellheim, S.; Guda, C.; Jansen, R. K.; Luo, H.; Tomkins, J.; Rognli, O. A.; Daniell, H. and Clarke, J. L. (2007): Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes, Theoretical and Applied Genetics 115 [4], pp. 571-90.

Sato, N.; Albrieux, C.; Joyard, J.; Douce, R. and Kuroiwa, T. (1993): Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids, EMBO (European Molecular Biology Organization) Journal 12 [2], pp. 555-61.

Satoh, J.; Baba, K.; Nakahira, Y.; Tsunoyama, Y.; Shiina, T. and Toyoshima, Y. (1999): Developmental stage-specific multi-subunit plastid RNA polymerases (PEP) in wheat, The Plant Journal 18 [4], pp. 407-15.

Scharff, L. B.; Childs, L.; Walther, D. and Bock, R. (2011): Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites, PLoS Genet 7 [6], p. e1002155.

Schein, A.; Sheffy-Levin, S.; Glaser, F. and Schuster, G. (2008): The RNase E/G-type endoribonuclease of higher plants is located in the chloroplast and cleaves RNA similarly to the *E. coli* enzyme, RNA 14 [6], pp. 1057-68.

Schiffer, S.; Rosch, S. and Marchfelder, A. (2002): Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes, EMBO (European Molecular Biology Organization) Journal 21 [11], pp. 2769-77.

Schimper, A. F. W. (1883): Über die Entwickelung der Chlorophyllkörner und Farbkörper Bot. Zeit. 41, pp. 105-114.

Schmitz-Linneweber, C. and Small, I. (2008): Pentatricopeptide repeat proteins: a socket set for organelle gene expression, Trends in Plant Science 13 [12], pp. 663-70.

Schmitz-Linneweber, Christian and Barkan, Alice (2007): RNA splicing and RNA editing in chloroplasts, Bock, Ralph, Cell and Molecular Biology of Plastids 19 pp. 213-248, Springer Berlin / Heidelberg.

Schweer, J.; Loschelder, H. and Link, G. (2006): A promoter switch that can rescue a plant sigma factor mutant, FEBS Letters 580 [28-29], pp. 6617-22.

Schweer, J.; Turkeri, H.; Kolpack, A. and Link, G. (2010): Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription - recent lessons from *Arabidopsis thaliana*, European Journal of Cell Biology 89 [12], pp. 940-6.

Serino, G. and Maliga, P. (1998): RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme, Plant Physiology 117 [4], pp. 1165-70.

Severinov, K.; Mustaev, A.; Kukarin, A.; Muzzin, O.; Bass, I.; Darst, S. A. and Goldfarb, A. (1996): Structural modules of the large subunits of RNA polymerase. Introducing archaebacterial and chloroplast split sites in the beta and beta' subunits of *Escherichia coli* RNA polymerase, Journal of Biological Chemistry 271 [44], pp. 27969-74.

Sexton, T. B.; Christopher, D. A. and Mullet, J. E. (1990): Light-induced switch in barley *psbD-psbC* promoter utilization: a novel mechanism regulating chloroplast gene expression, EMBO (European Molecular Biology Organization) Journal 9 [13], pp. 4485-94.

Sexton, T. B.; Jones, J. T. and Mullet, J. E. (1990): Sequence and transcriptional analysis of the barley ctDNA region upstream of *psbD-psbC* encoding *trnK*(UUU), *rps16*, *trnQ*(UUG), *psbK*, *psbI*, and *trnS*(GCU), Current Genetics 17 [5], pp. 445-54.

Shanklin, J.; DeWitt, N. D. and Flanagan, J. M. (1995): The stroma of higher plant plastids contain ClpP and ClpC, functional homologs of *Escherichia coli* ClpP and ClpA: an archetypal two-component ATP-dependent protease, Plant Cell 7 [10], pp. 1713-22.

Sharma, C. M.; Hoffmann, S.; Darfeuille, F.; Reignier, J.; Findeiss, S.; Sittka, A.; Chabas, S.; Reiche, K.; Hackermuller, J.; Reinhardt, R.; Stadler, P. F. and Vogel, J. (2010): The primary transcriptome of the major human pathogen *Helicobacter pylori*, Nature 464 [7286], pp. 250-5.

Sharwood, R. E.; Halpert, M.; Luro, S.; Schuster, G. and Stern, D. B. (2011): Chloroplast RNase J compensates for inefficient transcription termination by removal of antisense RNA, RNA 17 [12], pp. 2165-76.

Sharwood, R. E.; Hotto, A. M.; Bollenbach, T. J. and Stern, D. B. (2011): Overaccumulation of the chloroplast antisense RNA AS5 is correlated with decreased abundance of 5S rRNA in vivo and inefficient 5S rRNA maturation in vitro, RNA 17 [2], pp. 230-43.

Shimada, H. and Sugiura, M. (1991): Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes, Nucleic Acids Research 19 [5], pp. 983-95.

Shinozaki, K.; Ohme, M.; Tanaka, M.; Wakasugi, T.; Hayashida, N.; Matsubayashi, T.; Zaita, N.; Chunwongse, J.; Obokata, J.; Yamaguchi-Shinozaki, K.; Ohto, C.; Torazawa, K.; Meng, B. Y.; Sugita, M.; Deno, H.; Kamogashira, T.; Yamada, K.; Kusuda, J.; Takaiwa, F.; Kato, A.; Tohdoh, N.; Shimada, H. and Sugiura, M. (1986): The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression, EMBO (European Molecular Biology Organization) Journal 5 [9], pp. 2043-2049.

Siemenroth, A.; Wollgiehn, R.; Neumann, D. and Börner, T. (1981): Synthesis of ribosomal RNA in ribosome-deficient plastids of the mutant “albostrians” of *Hordeum vulgare* L., Planta 153 [6], pp. 547-555.

Sijben-Muller, G.; Hallick, R. B.; Alt, J.; Westhoff, P. and Herrmann, R. G. (1986): Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase alpha-subunit, Nucleic Acids Research 14 [2], pp. 1029-44.

Silhavy, D. and Maliga, P. (1998): Mapping of promoters for the nucleus-encoded plastid RNA polymerase (NEP) in the *iojap* maize mutant, Current Genetics 33 [5], pp. 340-4.

Sittka, A.; Lucchini, S.; Papenfort, K.; Sharma, C. M.; Rolle, K.; Binnewies, T. T.; Hinton, J. C. and Vogel, J. (2008): Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq, PLoS Genet 4 [8], p. e1000163.

Small, I. D. and Peeters, N. (2000): The PPR motif - a TPR-related motif prevalent in plant organellar proteins, Trends in Biochemical Sciences 25 [2], pp. 46-7.

Smith, C.; Heyne, S.; Richter, A. S.; Will, S. and Backofen, R. (2010): Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA, Nucleic Acids Research 38 [Web Server issue], pp. W373-7.

Sriraman, P.; Silhavy, D. and Maliga, P. (1998): The phage-type P*clpP*-53 plastid promoter comprises sequences downstream of the transcription initiation site, Nucleic Acids Research 26 [21], pp. 4874-9.

Stern, D. B.; Goldschmidt-Clermont, M. and Hanson, M. R. (2010): Chloroplast RNA metabolism, Annual Review of Plant Biology 61, pp. 125-55.

Stocking, C. R. and Gifford Jr, E. M. (1959): Incorporation of thymidine into chloroplasts of Spirogyra, Biochemical and Biophysical Research Communications 1 [3], pp. 159-164.

Stoppel, R. and Meurer, J. (2011): The cutting crew - ribonucleases are key players in the control of plastid gene expression, J Exp Bot doi: 10.1093/jxb/err401.

Storz, G.; Altuvia, S. and Wassarman, K. M. (2005): An abundance of RNA regulators, Annu Rev Biochem 74, pp. 199-217.

Stothard, Paul and Wishart, David S. (2005): Circular genome visualization and exploration using CGView, Bioinformatics 21 [4], pp. 537-539.

Strittmatter, G.; Gozdzicka-Jozefiak, A. and Kossel, H. (1985): Identification of an rRNA operon promoter from *Zea mays* chloroplasts which excludes the proximal *tRNAVal* from the primary transcript, EMBO (European Molecular Biology Organization) Journal 4 [3], pp. 599-604.

Strittmatter, P.; Soll, J. and Bolter, B. (2010): The chloroplast protein import machinery: a review, Methods in Molecular Biology 619, pp. 307-21.

Sugiura, M. (1992): The chloroplast genome, Plant Molecular Biology 19 [1], pp. 149-68.

Sun, E.; Wu, B. W. and Tewari, K. K. (1989): In vitro analysis of the pea chloroplast 16S rRNA gene promoter, Molecular and Cellular Biology 9 [12], pp. 5650-9.

Surzycki, S. J. (1969): Genetic functions of the chloroplast of *Chlamydomonas reinhardi*: effect of rifampin on chloroplast DNA-dependent RNA polymerase, Proceedings of the National Academy of Sciences of the United States of America 63 [4], pp. 1327-34.

Suzuki, J. Y. and Maliga, P. (2000): Engineering of the *rpl23* gene cluster to replace the plastid RNA polymerase alpha subunit with the *Escherichia coli* homologue, Current Genetics 38 [4], pp. 218-25.

Swiatecka-Hagenbruch, M.; Emanuel, C.; Hedtke, B.; Liere, K. and Börner, T. (2008): Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase, Nucleic Acids Research 36 [3], pp. 785-92.

Swiatecka-Hagenbruch, Monika; Liere, Karsten and Börner, Thomas (2007): High diversity of plastidial promoters in *Arabidopsis thaliana*, Molecular Genetics and Genomics 277 [6], pp. 725-734.

Tanaka, K.; Oikawa, K.; Ohta, N.; Kuroiwa, H.; Kuroiwa, T. and Takahashi, H. (1996): Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga, Science 272 [5270], pp. 1932-5.

Tanaka, K.; Tozawa, Y.; Mochizuki, N.; Shinozaki, K.; Nagatani, A.; Wakasa, K. and Takahashi, H. (1997): Characterization of three cDNA species encoding plastid RNA polymerase sigma factors in *Arabidopsis thaliana*: evidence for the sigma factor heterogeneity in higher plant plastids, FEBS Letters 413 [2], pp. 309-13.

Thomas, B. C.; Li, X. and Gegenheimer, P. (2000): Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism, RNA 6 [4], pp. 545-53.

Vaistij, F. E.; Boudreau, E.; Lemaire, S. D.; Goldschmidt-Clermont, M. and Rochaix, J. D. (2000): Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast psbB/psbT/psbH gene cluster in Chlamydomonas reinhardtii, Proceedings of the National Academy of Sciences of the United States of America 97 [26], pp. 14813-8.

Vera, A.; Hirose, T. and Sugiura, M. (1996): A ribosomal protein gene (*rpl32*) from tobacco chloroplast DNA is transcribed from alternative promoters: similarities in promoter region organization in plastid housekeeping genes, Molecular and General Genetics 251 [5], pp. 518-25.

Vera, A. and Sugiura, M. (1995): Chloroplast rRNA transcription from structurally different tandem promoters: an additional novel-type promoter, Current Genetics 27 [3], pp. 280-4.

Vogel, J. and Wagner, E. G. (2007): Target identification of small noncoding RNAs in bacteria, Current Opinion in Microbiology 10 [3], pp. 262-70.

Wallin, Ivan E (1927): Symbionticism and the Origin of Species, The Williams & Wilkins Company, Baltimore.

Walter, M.; Kilian, J. and Kudla, J. (2002): PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts, EMBO (European Molecular Biology Organization) Journal 21 [24], pp. 6905-14.

Walter, Michael; Piepenburg, Katrin; Schöttler, Mark Aurel; Petersen, Kerstin; Kahlau, Sabine; Tiller, Nadine; Drechsel, Oliver; Weingartner, Magdalena; Kudla, Jörg and Bock, Ralph (2010): Knockout of the plastid RNase E leads to defective RNA processing and chloroplast ribosome deficiency, The Plant Journal 64 [5], pp. 851-863.

Wang, L.; Yu, X.; Wang, H.; Lu, Y. Z.; de Ruiter, M.; Prins, M. and He, Y. K. (2011): A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*), BMC Genomics 12, p. 289.

Wang, M. J.; Davis, N. W. and Gegenheimer, P. (1988): Novel mechanisms for maturation of chloroplast transfer RNA precursors, EMBO (European Molecular Biology Organization) Journal 7 [6], pp. 1567-74.

Weihe, A. and Börner, T. (1999): Transcription and the architecture of promoters in chloroplasts, Trends in Plant Science 4 [5], pp. 169-170.

Westhoff, Peter (1985): Transcription of the gene encoding the 51 kd chlorophyll *a*-apoprotein of the photosystem II reaction centre from spinach, Molecular and General Genetics 201 [1], pp. 115-123.

Williams-Carrier, R.; Kroeger, T. and Barkan, A. (2008): Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand, RNA 14 [9], pp. 1930-41.

Winkler, W. C. and Breaker, R. R. (2005): Regulation of bacterial gene expression by riboswitches, Annu Rev Microbiol 59, pp. 487-517.

Wu, Chih-Yu; Lin, Chi-Hui and Chen, Liang-Jwu (1997): Identification of the transcription start site for the spinach chloroplast serine tRNA gene, FEBS Letters 418 [1-2], pp. 157-161.

Xie, G. and Allison, L. A. (2002): Sequences upstream of the YRTA core region are essential for transcription of the tobacco *atpB* NEP promoter in chloroplasts in vivo, Current Genetics 41 [3], pp. 176-82.

Yamazaki, H.; Tasaka, M. and Shikanai, T. (2004): PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*, The Plant Journal 38 [1], pp. 152-63.

Yehudai-Resheff, S.; Hirsh, M. and Schuster, G. (2001): Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts, Molecular and Cellular Biology 21 [16], pp. 5408-16.

Zghidi-Abouzid, O.; Merendino, L.; Buhr, F.; Malik Ghulam, M. and Lerbs-Mache, S. (2011): Characterization of plastid *psbT* sense and antisense RNAs, Nucleic Acids Research 39 [13], pp. 5379-87.

Zghidi, Wafa; Merendino, Livia; Cottet, Annick; Mache, Regis and Lerbs-Mache, Silva (2006): Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the *psbN* gene in plastids, Nucleic Acids Research 35 [2], pp. 455-464.

Zhelyazkova, P.; Hammani, K.; Rojas, M.; Voelker, R.; Vargas-Suarez, M.; Borner, T. and Barkan, A. (2011): Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts, Nucleic Acids Research doi: 10.1093/nar/gkr1137

Zoschke, R.; Liere, K. and Börner, T. (2007): From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development, The Plant Journal 50 [4], pp. 710-22.

Zubo, Y. O.; Yamburenko, M. V.; Selivankina, S. Y.; Shakirova, F. M.; Avalbaev, A. M.; Kudryakova, N. V.; Zubkova, N. K.; Liere, K.; Kulaeva, O. N.; Kusnetsov, V. V. and Börner, T. (2008): Cytokinin stimulates chloroplast transcription in detached barley leaves, Plant Physiology 148 [2], pp. 1082-93.

Zuker, Michael and Stiegler, Patrick (1981): Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, Nucleic Acids Research 9 [1], pp. 133-148.

6 APPENDICES

Abbreviations

%	percentage
°C	degree Celsius
5'-P	5'-monophosphate
5'-PPP	5'-triphosphate
A, T, G, C, U	nucleic acid bases (adenosine, thymine, guanine, cytosine,uracil)
asRNA	antisense RNA
ATP	adenosine-5'-triphosphate
ATPase	adenosine triphosphatase
aTSS	antisense TSS
B. subtilis	*Bacillus subtilis*
BLRP	blue-light-responsive promoter
cDNA	complementary DNA
cRT-PCR	circular reverse transcription-PCR
DNA	deoxyribonucleic acid
dRNA-seq	differential RNA sequencing
E.coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
gTSS	gene TSS
IR	inverted repeat
iTSS	internal TSS
kb	kilobase
kPa	kilopascal
LSC	large single copy
M	molar
MAST	Multiple Alignement Searching Tool
MEME	Multiple Expectation-Maximization for Motif Elicitation
$\mu E\ s^{-1}\ m^{-2}$	micoreinstein per second per square meter
MFE	minumum free energy
µg	microgram
min	minute
mL	milliliter
µL	microliter
mM	milimolar
mRNA	messenger RNA
ncRNA	non-coding RNA
NEP	nuclear-encoded plastid RNA polymerase
nm	nanometer

nt	nucleotide
ORF	open reading frame
oTSS	orphan TSS
PCR	polymerase chain reaction
PEP	plastid-encoded plastid RNA polymerase
pH	potential hydrogen, -log [H+]
PNPase	polynucleotide phosphorylase
PPR	pentatricopeptide repeat
PS	processing site; 5'-P end
PSI	photosystem I
PSII	photosystem II
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
RNA-seq	RNA sequencing; transcriptome sequencing
rpm	round per minute
RpoT	RNA polymerase of phage T3/T7 type
RpoTm	mitochondrial RpoT
RpoTmp	mitochondrial and plastid RpoT
RpoTp	plastid RpoT
rRNA	ribosomal RNA
RT	reverse transcription
RuBisCO	ribulose 1,5-bis-phosphate carboxylase/oxygenase
sRNA	small RNA
SSC	small single copy
TAE	Tris-acetate EDTA buffer
TEX	terminator exonuclease
tRNA	transfer RNA
TSS	transcription start site; 5'-PPP end
U	unit
UTR	untranslated region
v/v	volume per total volume

Appendix A: List of TSSs revealed by dRNA-seq of barley plastids. A-1. TSSs mapped in green plastids. 2. TSSs mapped in white plastids. The tables provide the TSSs mapped in green and white plastids by dRNA-seq. The TSSs are marked with a T and named after the downstream located gene and the number of nucleotides between the 5'-PPP end and the start codon of the ORF. The TSS categories and the associated genes are given. The 50 nt upstream sequence and the abundance (number of cDNAs in +/- libraries) of each 5'-PPP end are provided. gTSSs, which are represented by cDNAs that do not reach the downstream located gene, are referred to as "disconnected" in the Comments column. 5' ends mapped within five consecutive nucleotides were considered to be a single TSS and denoted by the genomic position of the most abundant of the 5' ends in the (+) library. The less abundant 5' ends are referred to as minor 5' ends and listed in the Comments column. TSSs found not to be enriched in green libraries after the TEX treatment are marked. Predicted PEP and NEP promoters are underlined and highlighted, respectively

Appendix A-1: TSSs mapped in green plastids

TSS	Genomic position	Strand	Gene TSS	Internal TSS	Antisense TSS	Orphan TSS	Not enriched	Sequence -50 nt upstream + TSS (51nt)	cDNAs (+/-)	Comments
TpsbK-6580	687	+			*psbA*			AACTTCAACAGCAGCTAAGTCTAGAGGGAAGTTGTGAGCATTACGTTCGTG	3/0	
TpsbA-80	1760	-	*psbA*					TGGGCTGACTTGGTTGACATTGGTATATAGTCTATGTTATACTGTTAAATA	7938/ 1235	Boyer and Mullet, 1988
TmatK-295	4036	-	*matK*	*trnK-UUU*				TTTTTTTTTAGAAATCCTAATTATTCTTGATTATGGATTGAATAAGGGATG	2/0	disconnected
TtrnK-239	4707	-	*trnK-UUU*					ACTTCGCTCAAATGATAAGGGTGTTCCTCTTGCATGTATTCTCATACAATA	2/1	disconnected
TtrnK-337	4805	-	*trnK-UUU*					ACCATTGTATTTCCATTGACAAAGGTCTATTGAACAAATAGAATTTGTAGA	8/2	disconnected; minor 5' ends: 4805,4806
TpsbK-1093	6174	+			*rps16*			TATGGATGGAATCATGAATAGTCATTGGTTTTGTTTTTTGTATACTAATTA	3/0	
TpsbK-783	6484	+				x	x	CTCCAAAAAAGTTTAATTCATTTAATTACTAGAATTAGAATTCTATTAGTA	2/3	TSS in white

TtrnQ-21	6939	-	*trnQ-UUG*					GTAATTGAAATGGTATTGACGAATAACAAATAAACATAATAGTGTTTATTA	6/0	
TtrnQ-150	7068	-	*trnQ-UUG*					AGAGTGGAATAAATTCGACTTTCATATAATATCTTAATAGAATTCCATGTA	10/0	disconnected
TpsbK-171	7096	+	*psbK*					ATCAAAAAAATGCATTGATCATTACATGGAATTCTATTAAGATATTATATG	28/22	Sexton, et al., 1990; Sexton, et al., 1990; disconnected; minor 5' ends: 7094,7096
TtrnS+1	8177	-	*trnS-GCU*					TCTTATTTCTTGCCTATATCATATCACGGAAACCTTTCGCTTTGGAACGTG	434/78	TSS at +1 of *trn* gene
TtrnS-2	8179	-	*trnS-GCU*					TTTCTTATTTCTTGCCTATATCATATCACGGAAACCTTTCGCTTTGGAACG	4/0	
TpsbD-716	8443	+	*psbD*					GATTTTCTCTTTAGCGGGCATTTCCATATAGGACTTGTTATAATTATAATA	2/0	disconnected
TpsbD-711	8448	+	*psbD*					TCTCTTTAGCGGGCATTTCCATATAGGACTTGTTATAATTATAATAAAACA	4/2	Sexton, et al., 1990; Sexton, et al., 1990; disconnected
TpsbD-557	8602	+	*psbD*				x	GTAAGTAGACCTGACTCCTTGAATGATGCCTCTATCCGCTATTCTGATATA	32/58	Sexton, et al., 1990; Sexton, et al., 1990; disconnected; minor 5' ends: 8602,8604
TtrnS-485	8662	-	*trnS-GCU*					TAGTAACAAGAATATGAAATCGTAAATAGCGAAAAATTCTTGTCTTGCGCG	18/0	disconnected
TtrnS-583	8760	-	*trnS-GCU*					CTAAAATGGAGGATTCTGAAAAAAAAAAGGACCTTCGAAATCAATTTTTAT	2/0	disconnected
TpsbC-194	9974	+	*psbC*	*psbD*				CAAATCTTTGGTGTTGCTTTTTCCAATAAACGTTGGTTACATTTCTTTATG	17/10	Sexton, et al., 1990; Sexton, et al., 1990; minor 5' ends: 9972, 9974
TtrnS-1984	10161	-			*psbC; psbD*			GTTTCCACGTGGTAGAACCTCCTCAGGGAATATAAGATTTTCATGAGGCTG	3/0	
TtrnS+1	11822	-	*trnS-UGA*					TTTTATATGGCTATGTTCTATTTGTAGGAGTAAAATAAGGATTAGGCGGTG	55/22	TSS at +1 of *trn* gene
TtrnS-7	11829	-	*trnS-UGA*					TTTTATTTTTTATATGGCTATGTTCTATTTGTAGGAGTAAAATAAGGATTA	1/0	TSS of a *trn* precursor
TtrnG-90	12557	+	*trnG-GCC*					CATTAGAATATTCATTGACAGATAATAAAAAAAGGAAAACTCTAATATCTA	84/29	
TtrnT-2278	12925	+				x		GGAGGCAGAAATAGAACTAGCCAACAGAGCCCCAAAATTGGGTATTAACTG	2/0	

TtrnfM+1	13239	-	*trnfM-CAU*					TAAAGGCTTGTATTCAAGCCTTTTTTGTCCACCAGTTTCTGGTACTACAGA	330/90	TSS at +1 of *trn* gene
TtrnT-1457	13746	+			*trnG-UCC*			AGCTATTTGGCTTCCTTATCCTTTAACAAAAGAAGATTTAGTTACGATTGG	2/0	
TtrnT-1326	13877	+			*trnG-UCC*			TATGCTTCGCGACTCTGTACTCATATCATAATCCAATTTTTATTTTGGATG	4/0	
TtrnG-3	14084	-	*trnG-UCC*					TCAAATGGCTATCCTTGACAAAGGGTACCATTTATACCATAATATACAATG	93/92	
TtrnT-844	14359	+				x		TTATTTTGATTAGGATCCTATCCTATTCTACGGTTACGACTACAATAATCA	3/0	
TtrnG-1036	15117	-			*trnT-GGU*			CCCATTATGGGTCTACTGCATAATGTACATATTATATATATATAATATATA	7/2	minor 5' ends: 15115, 15117, 15121
TtrnT-70	15133	+	*trnT-GGU*					TACAATCTATACATAATGTCTCTCTCTATATCTCTATATATTATATATATA	35/10	minor 5' ends: 15131,15133
TtrnT-59	15144	+	*trnT-GGU*					CATAATGTCTCTCTCTATATCTCTATATATTATATATATATAATATGTACA	26/6	
TtrnG-1659	15740	-			*trnE-UUC*			ATCACTAGACGATAGGGCCATATACAACCGCTCGTGATTATACTATAATCA	8/1	
TtrnE-46	15745	+	*trnY-GUA; trnE-UUC*					ATACTTTTCGAGTATATTGACAATTCAAAAAAACTGCTCATACTATGATTA	144/56	
TtrnE+1	15791	+	*trnE-UUC*					GATTATAGTATAATCACGAGCGGTTGTATATGGCCCTATCGTCTAGTGATG	729/29	TSS at +1 of *trn* gene
TtrnG-1808	15889	-			*trnE-UUC; trnY-GUA*			CTCGGGCATCGACCCAGGAAGAATTTATTCTAGGGTTTTTGCTAATCTATG	5/1	minor 5' ends: 15885, 15889
TtrnD-133	16232	+	*trnD-GUC*					TTATCCATTTTTAGTGATAAAAAATCACGACATACTAGTTATGTCACTCTC	5/0	disconnected
TtrnG-2782	16863	-				x		TCTTATACAATTTCTCTTCCTTTTTCTTATAGTTATACATACAATTATGTA	4/0	
TpsbM-348	16868	+	*psbM*				x	ATTTGGATGTCTATGTGACCCATAGAAAGTTGCTCATATAATACATACATA	0/7	TSS in white
TpsbM-162	17054	+	*psbM*					CCATTTTGGTCTTAGTATGGATATAGGATCTTCCTATGTTATACTATATAA	15/4	minor 5' ends: 17054-17057
TpsbM-111	17105	+	*psbM*					CTCTCATCCATGAATTGATTTGATAGATCCGATATTCATAATATTGAATTG	71/24	
TrpoB-2586	17501	+				x		ACTTTAGAGTCTTTCTATTATATTATCTTTGAATCTACATAGAATAGATTA	3/1	
TtrnG-3435	17516	-				x	x	ATTGAAATTAAGTGGATGCAATGAAAAAAGAAATTGAATTAGACTACTTAA	7/8	

TrpoB-2539	17548	+				x		ATTAGTAGATTGAAATAGTTAAGTAGTCTAATTCAATTTCTTTTTTCATTG	47/10	
TtrnG-3476	17557	-				x		GGTCAAGAATTGTCTTCGATTTTTTTATTTTGACCTGATTGAAATTAAGTG	21/0	minor 5' ends:17553, 17557;
TpetN-44	18170	-	*petN*				x	CAGAAGCCCTTTTTGACTCTGTACCCTTGATTCCACTATGATTATTGATCA	159/ 143	
TtrnC-377	19298	-	*trnC-GCA*					CCTAAAGTTCACCTATACACGGATTAAGCATAAAGCCATAATATTTTATTA	17/3	disconnected; minor 5' ends: 19298, 19301
TtrnC-808	19729	-				x		TTCTCATTTTATCCCCTTTCTATGAATGTAATAAAGCCGTAATATAATACA	2/0	
TrpoB-147	19940	+	*rpoB*					AGAATTTTCGTCGAAATGGTCTCTATTCATATGTATGAAATACATATATGA	12/9	Liere and Börner, 2007; Silhavy and Maliga, 1998
TtrnC-5056	23977	-			*rpoC1*		x	CGTCTTCTAATTCCTTCCATTCGACCAAAGAATTCTCTATAATAATTCGCA	3/3	
Trps2-1914	28459	+		*rpoC2*				CAAAGAATTTATCACACAAATAGTCAATTTGTTAGAACTTGCTTAGTAGTG	2/0	
TtrnC-9794	28715	-			*rpoC2*			AGAACAGTTGGATGCCGACAAAATCATCAAAGATTGGTATTCTTTATTTCG	3/0	
Trps2-261	30112	+	*rps2*					ATTTAACCCTAAGGATACATAAAACATATTTTTTACTTTACTAGACTTTTG	3/0	disconnected
Trps2-152	30221	+	*rps2*					AAACAAGTCAGTTAATTCATTAAATTAAGGTTTTGTTTATACCATGTATCA	17/10	disconnected; TSS in white; minor 5' ends: 30221, 30224
TatpI-140	31208	+	*atpI*					AATTTGTTTTGCATAGAAAAAAGAAGGGGAATATTGATATATATTAGAGGG	10/1	disconnected; minor 5' ends: 31204,31208
TatpI-118	31230	+	*atpI*					GAAGGGGAATATTGATATATATTAGAGGGTATTGATATATATTATGATCTG	96/46	
TtrnC-12426	31347	-			*atpI*			ATATCGTATAACCCCTTGAGTGTTTTAATGGAACAAGGTATAATATTCATA	11/2	minor 5' ends: 31347, 31349
TatpH-783	31876	+		*atpI*				ATTTCGAGAAATATATTAAACCAACCCCAATCCTTTTACCAATTAACATCC	3/0	
TatpH-350	32309	+	*atpH*					TATGCATTAAGGAGGGTGGAGTCAGGCTAGATCTATACTATAATAAATATC	2/0	disconnected
TtrnC-13514	32435	-				x		ATCGGATTAGATAATGAATCTAACCTAGGAATATATAATATAATATCAATA	3/0	
TatpH-201	32458	+	*atpH*					AAATAGAAGAAACAAATGTATATAGGATATTGATATTATATTATATTCC	9/0	disconnected; minor 5' ends: 32457,32458
TatpH-175	32484	+	*atpH*					ATATTGATATTATATTATATATTCCTAGGTTAGATTCATTATCTAATCCGA	6/0	disconnected
TatpF-388	32972	+	*atpF*					CTTTCGATTTCGATTAGATACTTTTTTCTTTTTTTTAGTAAATTGGTATTTG	2/0	disconnected
TtrnC-14756	33677	-			*atpF*			TTTAACCGTGTTAATGGTCTCACATTCTTGGTTTATAGAGAATCAAAATTG	5/1	
TatpA-897	33939	+		*atpF*				GGGCACCCTGCCCTTTTTTAACCAATGCCGAATCGACGACCTATGTATAAA	2/0	

TtrnC-15956	34877	-			*atpA*			CGACCTATATTCTCAATCCCTACTTTCCTATTATATTGTTCAATACGTTCG	2/0	
TtrnM-69	3682	-	*trnfM-CAU*					CATAGAGGAGCCCTCTTTACCATTCTGTATAAATGGACTATTCTATTTGTA	14/4	
TtrnS-7862	37295	+			*rps14; psaB*			TTGCCATAATGTGCCGTTGCTATTATTATCAAGTATACGGTTCTAATCCTA	2/0	
Trps14-733	37984	-	*rps14*	*psaB*				TGCTCTAGGTTTGCATACAACTACATTGATTTTAGTAAAGGGCGCTTTAGA	6/0	disconnected
TtrnS-4772	40385	+			*psaA*			CTACTGCTACTAACTCGCCGCCTCCCCACGTTAAGCTAGTACTTGTTGTTG	4/0	
TtrnS-4375	40782	+			*psaA*			TGTACCATAGTCAGTAGCTAGGTATGGATAGGGAGGCATAGAATACATATG	2/0	
TpsaB-1715	41317	-		*psaA*				TTTGCTGGTTGGTTCCATTATCACAAAGCCGCTCCCAAATTGGCCTGGTTC	2/0	
TpsaB-2144	41746	-		*psaA*				CCCGGCCATTTCTCAAGAACACTAGCTAAGGGCCCTGATACTACCACTTGG	2/0	
TtrnS-3158	41999	+			*psaA*			TTGAAAGATAGATTTTGAAAGATAGATTTTGAAAGATAGATATTGTGATTA	27/2	
TpsaA-209	42089	-	*psaA*				x	AAGATAAACAATGTCCGTTAGGCACCTAACCTTTATGTCATAATAGATCCG	263/369	Berends, et al., 1987; Swiatecka-Hagenbruch, et al., 2007
TtrnS-2690	42467	+			*ycf3*			ATGAAATATCTTATAAAATAGAAGGTATAAGAAATGGATATAATGAAATTC	2/0	
TtrnS-814	44343	+			*ycf3*			ACCTTCCCGAAGCGTGGGTTTGCTTGCTTTACAAATTTGGTTCTCTCTATG	2/0	
Tycf3-73	44619	-	*ycf3*					TAAAACATATGGATTTAGAGCATATGGATCAATCTATTTATAATTATAAGA	3/0	disconnected
TtrnS-5	45152	+	*trnS-GGA*					CAGGAATAAGAAAACTCGCTATTCACTCAGTTTATTTTCCATAATAAGTTA	12/5	minor 5' ends: 45152, 45154, 45156
Trps4-174	46324	-	*rps4*					ACCTTACTATTTTGCTAGATACAAAACAATAAAATAAATAATATATATACA	2/1	disconnected
TtrnT-9	46540	-	*trnT-UGU*					TTTTACGGATTTCCTATACTATACCTATTTCTATTTGTTACACTATTTCCG	11/4	
TtrnL-152	47003	+	*trnL-UAA*					ATATATTGATTCGGATTGAATTGCAAATACATCAAGGATAGAATCAATGTA	21/10	
TtrnF-938	47190	+		*trnL-UAA*				AAAGAAGAGGGAAGTGGGGATATGGCGAAATCGGTAGACGCTACGGACTTG	4/1	
TtrnT-1740	48271	-			*trnF-GAA*			TTTATCCACTTAGATGAATAAATCATACTCTATCTATATTATATATTATTA	12/3	
TtrnM-3791	48272	+				x		AATAGGTATTGATCCAAATACCTCGAGATGGATTGTGATACATATTTATTA	10/5	
TtrnM-3776	48287	+				x		AAATACCTCGAGATGGATTGTGATACATATTTATTAATAATATATAATATA	16/6	
TtrnT-1812	48343	-			*trnF-GAA*			TTCCTTATCATAAAGAAAGGGTATATTTTAAAACTTCTCTAAAAAGAAGTG	2/0	
TtrnT-2501	49032	-		*ndhJ*				CCGGTGGATCTTTAGCTAGCGTGTATCATCTTACAAGAATACAATACGGTA	4/3	

TndhJ-2580	49111	-		*ndhJ*				GATTGGGATTCCATTGCTGTCATTTTATATGTATATGGTTATAATTATTTA	5/0	minor 5' ends: 49109, 49111
TndhK-158	50264	-	*ndhK*	*ndhC*				AACCCATGGGAGGGGCTTGGGTACAATTCCGAATACGCTATTATATGTTTG	9/4	disconnected
TndhC-336	50795	-	*ndhC*					AATTAGTAGTATTCTCATTTTTATTTAATAGTCTCTTATTATTATTAAATA	2/0	disconnected
TndhC-376	50835	-	*ndhC*					TAGATTAGTCCTACTATCTTAATTAGGTTCTAATTATTCTAATTAGTAGTA	2/0	disconnected
TtrnM-582	51481	+	*trnM-CAU*		*trnV-UAC*			TCAAAGGAGGATTACCTTGACTTAGGTCTGCCTCTGGTCTAAATTAAATCA	51/11	disconnected
TndhC-1172	51631	-		*trnV-UAC*				GTGCCAATTCAGGTGCCTAATCAAATAGAACCCTTATGGATTTGCTAGTTG	3/0	
TtrnV-4	51861	-	*trnV-UAC*					TAGAAAATATTCATCTTGACAAGAAATTCTCTATATGTTAAGATATCTCTG	52/6	
TtrnV-89	51946	-	*trnV-UAC*		*trnM-CAU*		x	ATACCGCTCTGTTATTGGTCTGTTATATACTAAATACTAAACTAAATACTA	11/21	
TtrnV-101	51958	-	*trnV-UAC*		*trnM-CAU*			CTTTTCTAAGCAATACCGCTCTGTTATTGGTCTGTTATATACTAAATACTA	9/6	
TtrnM-95	51968	+	*trnM-CAU*					GAGGATATCCCTTTGATCTGTATCTGTTTAGTATTTAGTTTAGTATTTAGT	4/3	disconnected; minor 5' ends: 51966, 51968
TtrnM-32	52031	+	*trnM-CAU*					AATAACAGAGCGGTATTGCTTAGAAAAGGGATTCAATATATAATCGATCGA	42/14	minor 5' ends: 52030-52033
TatpB-294	54450	-	*atpB*					ATAAAATTCTTTGTTGACAGCAATCTATGCTTCACAGTAGTATATATTTTG	155/43	disconnected
TrbcL-426	54513	+	*rbcL*					ACTGTGAAGCATAGATTGCTGTCAACAAAGAATTTTATTAGTATTTAGTTA	3/0	disconnected
TatpB-392	54548	-	*atpB*					AGGTATAGCGCAACCCAAATTAATCCCTAATCCTTATTTTACAAGTTCTTA	2/0	disconnected
Trpl23-1635	55013	+		*rbcL*				GCAGGTGTTGGATTTCAAGCTGGTGTTAAAGATTATAAATTGACTTACTAC	2/0	
TatpB-857	55013	-			*rbcL*			ACTCGGAATGCTGCCAAGATATCAGTATCCTTAGTTTCATACTCTGGGGTG	2/0	
TatpB-1656	55812	-			*rbcL*			ACCATGATTTTTCTGTCTATCAATAACTGCATGCATTGCACGGTGAATGTG	3/0	
Trpl23-614	56034	+		*rbcL*				TGCGCGATGATTTTATTGAAAAAGATCGTGCTCGCGGTATCTTTTTCACTC	5/0	
TatpB-2409	56565	-			*rbcL*			TTTTTATTTCTTTATTGTATTATACCTTAATATATATATATATTCTAGATA	29/3	
Trpl23-75	56573	+	*rpl23*					ATGATCATGAGACTTGACAAATCGAGATTCGTCTATTCTATATATCTAGAA	5/0	minor 5' ends: 56570,56573
TatpB-2468	56624	-			*rbcL*			TGTAAATACTGCGTATTTGATTCCATTATCATATGATACTACTATATTTTA	5/3	minor 5' ends: 56623, 56624
TpsaI-1262	56720	+		*rpl23*				TTACAGAAAAAAGTCTTCGTTTATTGGGAAAGAATCAATATACTTTTAATG	2/0	
TpsaI-70	57912	+	*psaI*					AACTAAAAAAATACAATAGTCAATATTCCTTATAATAGATATACTTAATTA	49/42	minor 5' ends: 57912,57914

Tycf4-301	58123	+	*ycf4*					AAAAAAAAAAGATTGTCTAGTATTTTTTAGTGTAAGTAATATAATATGGTA	2/0	disconnected
Tycf4-2	58422	+	*ycf4*					TAGCTTATTCTCTCAATTTCAATCGACCGCTGCTGGATTTAGTATATCTAA	3/0	
TatpB-5115	59271	-				x		TTGATACCCTGTATCTATTCTTCTTTTTTGTATATTAGTGGAATTTCAGTG	2/0	
TatpB-5129	59285	-				x		AAAACTCTATAGGTTTGATACCCTGTATCTATTCTTCTTTTTTGTATATTA	3/0	
TpetA-112	60259	+	*petA*					TTCCATTTTACCAAAATTCTTTCTATTTATACCTGCTTTAAGGTATTCATC	2/0	disconnected
TpetL-2041	62043	+			*psbJ*		x	GATCTCTTTTTCTTGTTGCTTCATAAGAGTGAATCGAATAGATTCAATTCG	7/8	
TpetL-1582	62502	+			*psbF; psbL*			ATAAACCCCAGTATAGACTGGTACGATTCAATTCAACATTTTGTTCATTCG	10/4	
TpsbE-140	63062	-	*psbE*					CAGGCAAAACCAGATTGCTGTGCCATAGGAAGGATAGCTATACTAATTCGG	372/ 65	
TpetL-93	63991	+	*petL*				x	TTTACACTTCTGTATCTCACTCTATCTTGTTTTTTAGTATTATCTAAAATA	1/12	verified by 5'-RACE
TpetG-39	64320	+	*petG*				x	TTTTCTTGGTCATTGAGATTCGTGGATAATTTAGAGTACTATTTAGGGATA	6/7	minor 5' ends: 64317, 64320; verified by 5'-RACE
TpsbE-1628	64550	-			*petG*			TACAGAGCGTGATTCCATCTATCTTATGTCGAAGCAGAGTAAAATAATTTA	18/11	
TpsbE-1633	64555	-			*petG*			AATCCTACAGAGCGTGATTCCATCTATCTTATGTCGAAGCAGAGTAAAATA	7/2	
TtrnP-21	64898	-	*trnP-UGG*					AAGAAAATTATGATGTGGAAAAGAAGACAGGAATTGTGTACAATGGCATTG	11/2	
TpsaJ-247	64970	+	*psaJ*		*trnP-UGG*			TGTCTTCTTTTCCACATCATAATTTTCTTATTCTTTCTCTATCTATATATA	2/0	disconnected
TtrnP-135	65012	-	*trnP-UGG*					TCTTTCTTTTATAGGCGAAATTCTATTGGATTTTAGCTGTATATAATTCCT	1/0	
TpsaJ-198	65019	+	*psaJ*					TAGAACCCTCGTGTCATTTCTTCTTTTTGGTCTCATATAATTAAGGATT	5/2	disconnected; minor 5' ends: 65015,65019
TpsaJ-38	65179	+	*psaJ*					TCGGAAAATTTCGTCGATTATGGTTCAGTTTATATATAGCATAAGTTTATG	2/0	
TtrnP-1928	66805	-			*rps18*		x	ATTAAGTGGTAGGAATCGACGAGCTGGATTACTTTCTTTATAATATATATA	48/46	
TtrnP-1937	66814	-			*rps18*		x	TAAATTAAGATTAAGTGGTAGGAATCGACGAGCTGGATTACTTTCTTTATA	8/8	TSS in white
Trpl20-334	67630	-	*rpl20*					GTATGGATAACTCTTATTGTGAAAAGAGCTATTTACTCTATAATGGATAGG	2/1	
TpsbB-176	69235	+	*psbB*					GGCCAAGTTACCCATTGCATATTGGCACTTATCGAGTATAGAATAGATCTG	91/23	
TpsbN-46	71434	-	*psbN*		*psbH; psbT*		x	ATTTAATCATTGGTGTTGACTTTGTATACTATTCCGTTGTAGTTGTAAATA	19/ 329	Zghidi, et al., 2006; verified by 5'-RACE

TpsbN-61	71449	-	*psbN*		*psbH*			GTAGCCATAAATTATATTTAATCATTGGTGTTGACTTTGTATACTATTCCG	5/0	
TpsbN-495	71883	-	*psbN*		*petB*			ATATCTGACGAATTAGAATCATCTTGATAGAGATAACAGGCCCTTTCTATG	41/9	disconnected
TpsbN-708	72096	-	*psbN*		*petB*			TGAATCAGTGTAATGACTCGACAAAATCGATAGAAAAAAAGTTGTCTTTTG	4/0	disconnected
TpsbN-908	72296	-			*petB*			CATTAGAGGCTTCTTGACTTTTCTCTTCCCTATTTTGGATTTTTTTTTATG	3/0	
TpsbN-1326	72714	-			*petB*			AAAAAGTCATAGCAAAACCGGTAGCGACTTGTACTAGAAAACAAGTAAGTG	4/1	
TpsbN-2112	73500	-			*petD*			CTAGGTAGGGGACTTACATGTTGATTGGATGCGGAGACACCTTTGGTTGTG	3/0	
TpsbN-2132	73520	-			*petD*			AGTGAACCAGCCTATCCTTCCTAGGTAGGGGACTTACATGTTGATTGGATG	2/0	
TpsbN-2355	73743	-			*petD*			CTACATTCTTCCTTAGATCCCTTCTTTTACTCCGATAGTATTGCGGATCAC	3/0	
TpsbN-3371	74759	-			*petD*			TTGCATCTTTGGTACAATCTATATTTTCGCGAAATGGATCATAATAAAATA	62/29	
Trps8-142	77775	-	*rps8*					GAATTGAATTTTACCAAAATAGTTTCATTAGCTCCTGAAGTATTATAAATA	2/0	disconnected
Trps8-281	77914	-	*rps8*	*rpl14*				AATTGTACGTACATGTAAAGAATTCAAATGCGAAGACGGTATAATAATCCG	2/0	disconnected
TtrnH-2693	78902	+				x		TTTTATTTTTATATTGAATTTGTTAAAACAGAAAAATATAATATATAGTCA	4/0	
Trpl16-98	79832	-	*rpl16*					AGAATAACTAATCTTGTCTTCGAATTCTGTCCATTGATAGAATAAAAAATG	12/10	disconnected
TtrnV-9666	82718	+			*rpl2*			GAACTCAATCACTTGCTGCCGTTACTCAACAGTTTTCTGTTGAGGTCTATC	2/0	
Trpl23-71	83580	-	*rpl23*					TCTCATCCATCATCCATACATAACGAATTGGTATGGTATATTCATACCATA	2/1	Hübschmann and Börner, 1998; TSS in white
TtrnI-1090	84847	-				x		CCATGAACTATAGATAGAGAAGAATCATTCTGAGCGAGTACATAAGAAGCG	2/0	
TtrnI-1694	85451	-				x		AAAAATTCTATATGTCTATTCTATCTATGATATTTCTATATATATAGAATA	4/0	
TtrnV-6424	85960	+			*trnL-CAA*			TCTTAAGACTACCCTTGCAGTAGGATACCATCTGCTTTGGGTTCTTTATTA	4/0	
TtrnL+1	86217	-	*trnL-CAA*					ATCACATTAGATAGATATCATATTCATGGAATACAATTCACTTTCAAGATG	5038/750	TSS at +1 of *trn* gene
TtrnL-23	86240	-	*trnL-CAA*					CACCTTTGTCATATATTCCATATATCACATTAGATAGATATCATATTCATG	5/0	
TtrnL-838	87055	-		*ndhB*				AGGCCTTCCTCCACTAGCAGGTTTCTTCGGAAAACTCTATCTATTCTGGTG	11/0	minor 5' ends: 87055, 87057
TtrnV-5213	87171	+			*ndhB*			CATAGGGCTAAAGAGAGAGCCAAAAAAGGATCTTTCATGTATAATCCTGCA	3/0	
TndhB-275	89309	-	*ndhB*				x	ATAGAGCTCTTGCACATTTTCGTTAATCCATGAACAGAATCTATGTATGTA	2/12	verified by 5'-RACE; TSS in white
Trps7-212	90016	-	*rps7*	*rps12*				ACAGGATTTCCGATCCTAGCGGGAAAAGGAGGGAAACGGATACTCAATTTA	4/0	disconnected
TtrnV-102	92282	+	*trnV*					TTCGATAGGACCGGTTGACAATTGAATCCAATTTTTCCCATTATTTGACTA	17/3	

TtrnV+1	92384	+	*trnV-GAC*					CGCCCCTTTGCCTTAGGATTCGTTAATTCTCTTTCTCGATGGGACGGGGAA	28/9	TSS at +1 of *trn* gene
Trrn16-116	92569	+	*rrn16*					AGAGGCTTGTGGGATTGACGTGATAGGGTAGGGTTGGCTATACTGCTGGTG	26/3	Hübschmann and Börner, 1998
TtrnI-608	93876	+	*trnI-GAU*	*rrn16*				CCAAGTCATCATGCCCCTTATGCCCTGGGCGACACACGTGCTACAATGGGC	39/8	disconnected;
Trps12-4898	95556	-			*trnA-UGC*			TTTGTCTCAGTAGAGTCTTTCAGTGGCATGTTTCAGTCCTCTTCCCCATTA	3/0	
Trps12-9264	99922	-			*rrn5*			AGACGATCTTTGCATTTTAGATTTGGATCTTTTTCTTATTTCAAAATAGTG	7/0	minor 5' ends: 99920, 99922
Trps12-9771	100429	-			*trnR-ACG*			GAATCCTACTTGGGGAGATTTTATTCATTCTTTAATGTAAGAATTTTAATG	3/0	
TtrnN-225	100759	-	*trnN-GUU*					CGATGAGAATTTCTTGACTTTCCATATATAGAAAGAGATAGACTATAAATG	4/1	
Trps15-393	101689	+	*rps15*				x	CGCTATTGGTTTGATACGAATAATGGCAATCGTTTCAGTATGTTAAGGATA	27/28	disconnected
Trps15-228	101854	+	*rps15*					CTAAATGATATCAATTAAATGGTGTATCAATTCCATAAATTGCATATAGCA	53/44	disconnected; TSS in white
Trpl32-1830	104013	+			*ndhF*			GGAATAAAGCAGCTTGATAAGAACCTATACCTAGAGCTAACATCATATAAC	2/0	
TtrnL-608	106139	+	*trnL-UAG*					CTTCTTTGAATGTACTTTTATGTGTCGAATTACTTCGGTACAATATTCTTA	15/7	potential ncRNA candidate?
TtrnL-23	106724	+	*trnL-UAG*					TGGACTTTCCCAATATCGACGATTCCCGAGATAATAGCTATTATTCTTTTA	8/4	
TtrnN-9189	108411	+			*ndhD*			CTCATTCCTGGTAAGGCAAGAGAAGCCATTGAAAAGCTACTAAACATGGTA	2/0	
TtrnN-8056	109544	+			*ndhD; psaC*			CGGCAAAACAACAAGTATTGTTAACCAAGGAAAATAACTCATGATAAAGTG	2/0	
TtrnN-8047	109553	+			*ndhD; psaC*			AACAAGTATTGTTAACCAAGGAAAATAACTCATGATAAAGTGATAAAGACA	9/4	minor 5' ends: 109551-109553
TpsaC-270	110176	-	*psaC*					ATTTGTAATTCATTGATATTGCAATATTCAAATTGCAATAATTTATATTGA	5/1	disconnected; minor 5' ends: 110177, 110176
TndhG-114	111550	-	*ndhG*					ATTTTTTCTTATTTTAAATCAGACTAGATTTTGATGATATAATATAGTTTA	2/0	
TndhI-79	112307	-	*ndhI*					AACCTCTTCTCAACTTGTTTCACTATAAATAAGATAATACAATAATAGATA	41/32	minor 5' ends: 112307, 112309
TndhI-99	112327	-	*ndhI*				x	CTGGGCAATCTATTATTAACAACCTCTTCTCAACTTGTTTCACTATAAATA	4/10	disconnected; TSS in white

Appendix A-2: TSSs mapped in white plastids

TSS	Genomic location	Strand	Gene TSS	Internal TSS	Antisense TSS	Orphan TSS	Sequence -50 nt upstream + TSS (51nt)	cDNAs (+/-)	Comments
TpsbK-6660	607	+			*psbA*		AAGATATTGGGTATTTTTGTCTTTTCTTTCTTCAAAAATTCT*TATA*TGTTA	2/0	
TpsbA-80	1760	-	*psbA*				TGGGCTGACTTGGTTGACATTGGTATATAGTCTATGTTATAC*TGTT*AAATA	141/9	
TpsbA-146	1826	-	*psbA*				ACTCACTTCATTTACAAATACAAAATTATTGGTTTGGTTAATT*TATT*TATA	4/0	disconnected
TpsbA-981	2661	-		*matK*			GTCCTTTGTTGGTAAGGAATCAAATGCTGGAGAATTCATTTCT*CATA*GATA	3/0	
TpsbA-1399	3079	-		*matK*			CAAAAAGAAAATAAAAGACTAGTTAAATTCCTATATAACTCT*TATG*TATCA	6/0	
TpsbA-1419	3099	-		*matK*			TGAAATCCATTCTTTTTTTTCAAAAAGAAAATAAAAGACTAGTTAAATTCC	2/0	
TpsbK-4030	3237	+			*matK*		GAGAAAGAATCGCAATAAATGCAAAGATGGAACATCTTGGATACGGTATTG	2/0	
TpsbA-1605	3285	-		*matK*			TATTTCCCTTTTTAGAAGACAAATTTTTGCATTTGGATTATC*TATC*ACATA	2/0	
TmatK-260	4001	-	*matK*	*trnK-UUU*			GATTGAATAAGGGATGTTATGGATTGAATGTGTAAAAGAAGCAG*TATA*TTG	2/0	disconnected
TtrnK-239	4707	-	*trnK-UUU*				ACTTCGCTCAAATGATAAGGGTGTTCCTCTTGCATGTATTCT*CATA*CAATA	3/0	disconnected
TtrnK-942	5410	-		*rps16*			TGAAGGATTTTGCCTTCGAAGTTTCTCAGGCAGAATTAGCCAACTTGAGTC	3/0	
Trps16-95	6169	-	*rps16*				ACTTTCTTTTATCCATATTGCTCCATAGATAGCAAGTTTTAATTTTAATTA	2/0	disconnected
Trps16-206	6280	-	*rps16*				TTTTAAGCAAACTTGTTTAGTCGTCTTTTTCTATGTTTCATT*CATA*TGATA	10/0	disconnected
TpsbK-783	6484	+				x	CTCCAAAAAAGTTTAATTCATTTAATTACTAGAATTAGAATTC*TATT*AGTA	39/3	
Trps16-438	6512	-	*rps16*				CAAAAACTTTCAATCTATTCATTGGAATAGTTTATTTAGTTAT*TATA*GTTA	4/0	disconnected
Trps16-735	6809	-	*rps16*				CCTTCCGTCCCAGATTGTTTCTGATGAAACAAACGGTGAAAT*CATA*TAGTA	4/0	disconnected
TpsbK-169	7098	+	*psbK*				CAAAAAAATGCATTGATCATTACATGGAATTCTATTAAGATAT*TATA*TGAA	2/0	minor 5' ends: 7097, 7098
TpsbI-178	7687	+	*psbI*				CACCTAGGTTGAGCAGGTACCTTTAGGTACCTACACAATACC*TATT*CAATA	2/0	
TtrnQ-806	7724	-				x	AAAATGCAAAAATGAATTCTGGAAAAAAATTGCAAGAATTCATTTATAATA	4/0	
TtrnS+1	8177	-	*trnS-GCU*				TCTTATTTCTTGCCTATATCATATCACGGAAACCTTTCGCTTTGGAACGTG	213/14	TSS at +1 of *trn* gene
TpsbD-216	8943	+	*psbD*				AAATAGGAGTCTTGGAATAATGCTGAATTCAAAGGTTTATTTC*TATA*AGTA	2/0	disconnected
TpsbD-68	9091	+	*psbD*				ATTGAAAAAGGGCATTGAACGAGAAAGAAATCGTCCACAGA*TATA*AAACTA	1/0	

TpsbC-93	10075	+	*psbC*	*psbD*			CTTGGCTCTGAACTTACGTGCCTATGACTTTGTTTCCCAGGAAATCCGTGC	2/0	
TpsbZ-451	11728	+	*psbZ*		*trnS-UGA*		AAAGATAAGGGACAGAATAAAAAAAAATGAAAGAAACAAACG*TATT*CAATA	2/0	disconnected
TtrnG-82	12565	+	*trnG-GCC*				TATTCATTGACAGATAATAAAAAAAGGAAAACTCTAATATCTAA*TATA*ATA	13/1	disconnected
TtrnS-798	12620	-			*trnG-GCC*		GTTTTACCATTGAACTACGCTCGCTACGGTATATATTTTATAG*CGTA*TATC	4/0	
TtrnS-806	12628	-			*trnG-GCC*		AAGGAGATGTTTTACCATTGAACTACGCTCGCTACGGTATA*TATT*TTATAG	2/0	
TtrnfM+1	13239	-	*trnfM-CAU*				TAAAGGCTTGTATTCAAGCCTTTTTTGTCCACCAGTTTCTGGTACTACAGA	162/9	TSS at +1 of *trn* gene; TSS of *trn* precursos
TtrnG-16	14097	-	*trnG-UCC*				ATTAGCCTCCTTGTCAAATGGCTATCCTTGACAAAGGGTACCATT*TATA*CC	1/0	
TtrnG-1040	15121	-			*trnT-GGU*		ATTTCCCATTATGGGTCTACTGCATAATGTACATATTATATA*TATA*TAATA	37/3	minor 5' ends: 15119, 15121, 15124
TtrnT-78	15125	+	*trnT-GGU*				AATACATATACAATCTATACATAATGTCTCTCTCTATATCTC*TATA*TATTA	9/1	disconnected
TtrnT-68	15135	+	*trnT-GGU*				CAATCTATACATAATGTCTCTCTCTATATCTCTATATATTATA*TATA*TATA	9/0	
TtrnG-1153	15234	-			*trnT-GGU*		TAATTTAAAAAGCCCTTTATCGGATTTGAACCGATGACTTATGCCTTACCA	3/0	
TtrnE-412	15379	+	*trnE-UUC*				TCTGATAGAGTACTTTTCTACTAAATTGATTCGCTTTTTTCTT*TGTT*TTTG	4/0	disconnected
TtrnG-1652	15733	-			*trnE-UUC*		GACGATAGGGCCATATACAACCGCTCGTGATTATACTATAAT*CATA*GTATG	13/3	
TtrnE-39	15752	+	*trnY-GUA; trnE-UUC*				TCGAGTATATTGACAATTCAAAAAAACTGCTCATACTATGAT*TATA*GTATA	1/0	TSS of *trn* precursor
TtrnE+1	15791	+	*trnE-UUC*				GATTATAGTATAATCACGAGCGGTTGTATATGGCCCTATCGTCTAGTGATG	444/17	TSS at +1 of *trn* gene
TtrnD-34	16331	+	*trnD-GUC*				GATTCGTCTATTTCTTAGAGTACGACAGACGAATCGAATCTTCT*TATG*GTT	2/0	
TtrnG-2476	16557	-			*trnD-GUC*		ATCGGGAGTAGTGAAAAAAATTCCTATACTCCTTCCCTCCCTCTTTAATGA	2/0	
TpsbM-356	16860	+	*psbM*				TACTGCTTATTTGGATGTCTATGTGACCCATAGAAAGTTGCT*CATA*TAATA	2/0	disconnected

TpsbM-348	16868	+	*psbM*				ATTTGGATGTCTATGTGACCCATAGAAAGTTGCTCATATAATA*CATA*CATA	11/4	disconnected; minor 5' ends: 16868, 16872
TpsbM-338	16878	+	*psbM*				CTATGTGACCCATAGAAAGTTGCTCATATAATACATACATAAT*TGTA*TGTA	3/0	disconnected
TpsbM-160	17056	+	*psbM*				TTTTGGTCTTAGTATGGATATAGGATCTTCCTATGTTATACTA*TATA*ACTC	3/0	
TrpoB-2795	17292	+		*psbM*			TGCTACTGCACTGTTCATTCTAATTCCTACTTCTTTTTTACT*TATT*ATTTA	7/0	
TtrnG-3367	17448	-				x	TCTATTCTATGTAGATTCAAAGATAATATAATAGAAAGACTCTAAAGTGTC	3/0	
TrpoB-2591	17496	+				x	ACGACACTTTAGAGTCTTTCTATTATATTATCTTTGAATCTA*CATA*GAATA	6/0	
TtrnG-3562	17643	-				x	TTTAAGCATGAAAGATTCGCTGGTTTTTTCCTTTTACTTTTTAC*TATA*GTC	2/0	
TtrnC+1	18921	-	*trnC-GCA*				AAAACGTGGATGAATACTTGCTGTTTTCTAGATTTTCGATCGGATTTTTTG	8/1	TSS at +1 of *trn* gene
TrpoB-345	19742	+	*rpoB*				ACGCTATTCTAAATCGAAGTAAAGTAAAACTTTCTAGTGTAT*TATA*TTACG	2/0	disconnected
TrpoB-275	19812	+	*rpoB*				AAGGGGATAAAATGAGAAATCTTTGCCATTCAATCTGATTAGAA*TATT*ATA	2/0	
TrpoB-270	19817	+	*rpoB*				GATAAAATGAGAAATCTTTGCCATTCAATCTGATTAGAATAT*TATA*AAGTA	7/0	
TrpoB-161	19926	+	*rpoB*				CCCCTGGTAAACTAGAATTTTCGTCGAAATGGTCTCTATTCA*TATG*TATGA	9/2	minor 5' ends: 19926, 19927
TrpoB-156	19931	+	*rpoB*				TGGTAAACTAGAATTTTCGTCGAAATGGTCTCTATTCATATG*TATG*AAATA	5/0	
TrpoB-147	19940	+	*rpoB*				AGAATTTTCGTCGAAATGGTCTCTATTCATATGTATGAAATA*CATA*TATGA	223/12	Liere and Börner, 2007; Silhavy and Maliga, 1998
TrpoC1-2836	20519	+		*rpoB*			CTGGTATTTACTACCGCTCGGAATTAGATCATAAGGGAATTTC*TATC*TACA	3/0	
TrpoC1-2275	21080	+		*rpoB*			GTTCGGTTGCGGATCTGTTACAAGATCAATTCGGATTGGCTCTTGGTCGTT	2/0	
TrpoC1-1380	21975	+		*rpoB*			AAGTATTCCTTTAGTTAACCACCGTCGCTCTAACAAAAATACT*TGTA*TGCA	2/0	
TrpoC1-1222	22133	+		*rpoB*			TAGCTTATATGCCATGGGAAGGTTACAATTTTGAAGACGCAGTACTAATTA	3/0	
TtrnC-3444	22365	-			*rpoB*		CCTCAAGGCCGAATATGGCCCGTAATAATCCAGCTTCTGCGA*TATA*TGATG	4/0	
TrpoC1-979	22376	+		*rpoB*			TTTTAGTAGGTAAATTAACGCCTCAGATAGCGAGCGAATCAT*CATA*TATCG	6/1	
TrpoC1-599	22756	+	*rpoC1*	*rpoB*			GAAAGCTCGCTCGGATTAGCGGGGGATCTGCTAAAGAAACAT*TATA*GAATA	23/3	disconnected
TrpoC2-2241	23372	+		*rpoC 1*			AATAAATATAAATTCTTTTCTTATTTCTATTTTATGATTGACCAA*TATA*AA	4/0	minor 5' ends: 23371, 23372
Trps2-4722	25651	+		*rpoC 2*			GGGTACTTATGTATGGCGGAACGGGCCAATCTGGTCTTTCATAATAAAGAG	2/0	
Trps2-4582	25791	+		*rpoC 2*			GACTCTGGGCTTCTATCAAGCCACTACTACATCGATTTCGTTAGGAATCGA	2/0	

Trps2-3554	26819	+		*rpoC 2*			GTTATATAGACTTGCATGTAACTATTCAGAGTCAAGATATTC*TATA*TAGTG	3/0	
Trps2-3356	27017	+		*rpoC 2*			AAATGCACTGGAGTACTGATGTTTACCATGCGCCCGAATATC*AATA*TGGTA	11/1	
TtrnC-8437	27358	-			*rpoC 2*		TTATCTTTTCTGTATCGGGGATCGTCAAAATAAGCAAAAATAG*TATT*TCTA	3/0	
Trps2-207	30166	+	*rps2*				TTAGAACACTAACAGGTTAAATTTTAGATTTTTAAGATTTT*TATT*TTAATA	5/0	disconnected;
Trps2-152	30221	+	*rps2*				AAACAAGTCAGTTAATTCATTAAATTAAGGTTTTGTTTATAC*CATG*TATCA	16/1	disconnected
TatpI-626	30722	+	*atpI*	*rps2*			GTTCAGTGGTATGTTAACGAATTGGTCGATTACGAAAACTAGACTTTCTCA	3/0	disconnected
TtrnC-12294	31215	-			*rps2; atpI*		ATTAATTTGATATTTAGGATACCAAGAAAGCACATCAGATCATAA*TATA*TA	4/0	
TatpI-123	31225	+	*atpI*				AAAAAGAAGGGGAATATTGATATATATTAGAGGGTATTGATA*TATA*TTATG	3/0	
TtrnC-13376	32297	-				x	AACCCGCATAAAAGAGGACTCCACTTATAGATATTAAGGATATTTATTATA	2/0	
TatpH-228	32431	+	*atpH*				AATAGAAAATGAGAAAATATGCAAACAAAATAGAAGAAACAAA*TGTA*TATA	4/0	disconnected
TatpH-212	32447	+	*atpH*				ATATGCAAACAAAATAGAAGAAACAAATGTATATAGGATATTGA*TATT*ATA	4/0	disconnected
TtrnC-16521	35442	-			*atpA*		CATGGCCCCCTCCTCATGGAAAGTAGTTACTACTTGAGCTACGGAGGATGC	3/0	
TtrnC-17348	36269	-			*atpA*		TGAATAGCTTCCTTCAAAAGGATTTCCGCTTGCTCGGTGAA*TGTC*TTGCTA	2/0	
TtrnC-17519	36440	-			*atpA*		ATCCTATTGGACGCAATCTTATTTCTATCTATTCTTTAAATAAC*TATA*CTA	4/0	
TtrnR+1	36547	-	*trnR-UCU*				GACAACAAATAAAAGAAAGAATAAAAGAAAGAATAAAATACGAAAGAAAAG	3/0	TSS at +1 of *trn*gene
TtrnS-8479	36678	+			*trnfM -CAU*		CTCTTTTAGGAAATCAAAAAATCCAGATACAAATGGATGATG*TATA*TATCA	2/0	
TtrnfM-63	36836	-	*trnfM-CAU*				GGAGCCCTCTTTACCATTCTGTATAAATGGACTATTCTATTTG*TATA*GATA	1/0	
Trps14-1883	39134	-		*psaA*			TTTTTCTATTATTTCTTTCTACGCTGTCCTTAATAGCGAGTTGGTTACATC	2/0	
TpsaB-1817	41419	-		*psaA*			ACCTCTGGTTTTTTTCAGCTTTGGCGAGCATCTGGAATAACTAGTGAATTA	2/0	
TpsaA-209	42089	-	*psaA*				AAGATAAACAATGTCCGTTAGGCACCTAACCTTTATGTCATAATAGATCCG	5/0	
TpsaA-832	42712	-		*ycf3*			AGGTAGGAAACTCTCAAGTACGGTTCTAAGGGAAGGAACTGCC*TATT*CCGA	2/0	
TpsaA-1113	42993	-		*ycf3*			TCTTATGGAAAAAAGTCTTTTTCAAGGATTCTATAGAGATTT*CATA*TATGA	3/0	
Tycf3-82	44628	-	*ycf3*				TTATCCATTTAAAACATATGGATTTAGAGCATATGGATCAATC*TATT*TATA	7/1	disconnected
TtrnS-522	44635	+	*trnS-GGA*		*ycf3*		TTCACAAAATAACATAAAAACAAACACATTCGATTCTTATAAT*TATA*AATA	2/0	disconnected
Tycf3-107	44653	-	*ycf3*				TGGATAAGAGAGATATGGATTTTCCGCTCAGCTCAAATTGTATCCTTTTCC	2/0	disconnected

Tycf3-270	44816	-	*ycf3*				TGAAATCAAAAATAGAGTATGAGACAATTAATAACTTTGAAAACACGAAAT	2/0	disconnected
Tycf3-332	44878	-	*ycf3*				CTTATTGCATAATAAAATAAAGAAACCTTAGAATTCCGTTTG*CATA*AGATA	2/0	disconnected
Tycf3-1239	45785	-		*rps4*			GGGGCCCGCCAATTAGTCAACCATAGACATATTTTAGTTAATGGC*CGTA*TA	3/0	
Tycf3-1283	45829	-		*rps4*			TGGATAATATCCTTTTTCGATTGGGTATGGCTTCAACCATTCCTGGGGCCC	2/0	
Trps4-178	46328	-	*rps4*				TTTTACCTTACTATTTTGCTAGATACAAAACAATAAAATAAATAA*TATA*TA	30/2	minor 5' ends: 46326, 46328
TtrnT-1115	47646	-				x	CCTCGTCCGATTAATCCACTTTCTTTTTAAAAGTTCTCAAAACTTTGAATT	2/0	
TtrnM-3781	48282	+				x	GATCCAAATACCTCGAGATGGATTGTGATACATATTTATTAATAA*TATA*TA	3/1	
TtrnT-2150	48681	-				x	AAGGAATATTTTTAGGTTCTTTTCCTAGATGAAACGGAATACC*TATT*TATA	15/2	
TtrnT-2339	48870	-		*ndhJ*			ATGATATGGTGGGAATCTCTTATGATAATCATCCGCGCCTTAAA*CGTA*TCC	2/0	
TtrnM-1754	50309	+			*ndhC*		GAGCGCAAACATATAATAGCGTATTCGGAATTGTACCCAAGCCCCTCCCAT	15/4	
TndhC-249	50708	-	*ndhC*				ATTCTAGAGAAAGCGAGACAAAGTGAAACGATAAAGTTTTCTT*TATA*TAAG	7/1	disconnected
TndhC-329	50788	-	*ndhC*				AGTATTCTCATTTTTATTTAATAGTCTCTTATTATTATTAAATAGTAAATA	2/0	disconnected
TndhC-336	50795	-	*ndhC*				AATTAGTAGTATTCTCATTTTTATTTAATAGTCTCTTATTAT*TATT*AAATA	4/0	disconnected
TndhC-639	51098	-	*ndhC*				GTTTTCTTTAGTTTAGTATAGTTTTCATTTTCGCATTAGTATT*TGTT*TATA	2/0	disconnected
TtrnV+1	51857	-	*trnV-UAC*				AAATATTCATCTTGACAAGAAATTCTCTATATGTTAAGATATCTCTGATAA	17/1	TSS at +1 of *trn* gene
TtrnV-111	51968	-	*trnV-UAC*		*trnM-CAU*		TATTGAATCCCTTTTCTAAGCAATACCGCTCTGTTATTGGTCTGT*TATA*TA	2/0	
TatpB-450	54606	-	*atpB*				TCGTTATTAAACCATGGATTTGATTTACCAAATCCATCATTAT*TGTA*TACT	3/0	disconnected
TrbcL-214	54725	+	*rbcL*				AACAACAACTCAATTCCTATCGAATTCCTATAGTGGAATTCC*TATA*GGATA	7/2	
TatpB-593	54749	-	*atpB*				AAGAAAATTGAGGGCATGCTTAGGTTAATGAATATGTTTCATT*CATA*TATA	11/0	Liere and Börner, 2007; Silhavy and Maliga, 1998; disconnected
TatpB-2415	56571	-			*rbcL*		ATTTTCTTTTTATTTCTTTATTGTATTATACCTTAATATATA*TATA*TATTC	10/0	
Trpl23-60	56588	+	*rpl23*				GACAAATCGAGATTCGTCTATTCTATATATCTAGAATATATA*TATA*TATTA	10/0	
TatpB-2481	56637	-			*rbcL; rpl23*		GAAGACTTTTTTCTGTAAATACTGCGTATTTGATTCCATTAT*CATA*TGATA	15/0	
Trpl23+1	56648	+	*rpl23*				ACAATAAAGAAATAAAAAGAAAATAATAAAATATAGTAGTAT*CATA*TGATA	10/1	
TatpB-2500	56656	-			*rbcL; rpl23*		TGATTCTTTCCCAATAAACGAAGACTTTTTTCTGTAAATACTG*CGTA*TTTG	4/0	

TatpB-3209	57365	-				x	ATGGATTATAGAAGAATAGTATTCAAAATACTGCTCTGGACG*CATA*TAGTA	7/0	minor 5' ends: 57364, 57365
TpsaI-83	57899	+	*psaI*				AGAATATTTTGCAAACTAAAAAAATACAATAGTCAATATTCCT*TATA*ATAG	5/0	
TpsaI-74	57908	+	*psaI*				TGCAAACTAAAAAAATACAATAGTCAATATTCCTTATAATAGA*TATA*CTTA	3/0	
Tycf4-125	58299	+	*ycf4*				TATTTTAATTGAATCAACGAATATTTTTTGAATAGAAAGTCAA*TGTA*TCTA	2/0	disconnected
TatpB-4263	58419	-			*ycf4*		AACCTTTTAGAAGTTCTACCCATATGTGTTCTGATCGCCAATT*CATA*TTAG	2/0	
Tycf4-3	58421	+	*ycf4*				TTAGCTTATTCTCTCAATTTCAATCGACCGCTGCTGGATTTAG*TATA*TCTA	3/1	
TcemA-85	59366	+	*cemA*				TATCAAACCTATAGAGTTTTTGCTTCAAAGAAATAGAAATAT*CATG*AAATA	2/0	disconnected
TcemA-69	59382	+	*cemA*				TTTTTGCTTCAAAGAAATAGAAATATCATGAAATAGAAATATCAT*CATA*TA	3/0	disconnected
TpetA-568	59803	+		*cemA*			GGATCATCTCCATATCATTTTGCATTTCTCGACAAATATAATCTGTTTGGC	2/0	
TatpB-6318	60474	-			*petA*		CAGTTGCTTCTCGTGGGTTTTCATAACCCTGCTGCGCAAAAATGG*GATA*TG	2/0	
TpetL-3023	61061	+		*petA*			TAAGAAAAATACTACGTAAAGAAAAAGGGGGGTATGAAATATC*CATA*GTTG	63/4	
TpetL-2033	62051	+			*psbJ*		TTTCTTGTTGCTTCATAAGAGTGAATCGAATAGATTCAATTCG*CGTT*ATAA	11/0	
TpetL-935	63149	+				x	TCTGGTTTTGCCTGGTTCCGAAACATAATTACTTTTTTTTGTC*TATA*GGTA	2/0	
TpetL-761	63323	+				x	AATGATCAATAGGTTTGGATAAATAATTTAGGAAGGATATTCT*CATA*CTGA	4/0	
TpetL-444	63640	+	*petL*				AGACCCTCGGGAGGTCGTGGAATGCTTTTCTTCTCCTCTTATTC*CATA*TAG	7/2	disconnected; minor 5' ends: 63640, 63641
TpetL-108	63976	+	*petL*				ATGAATCCACTTTCTTTTACACTTCTGTATCTCACTCTATCT*TGTT*TTTTA	4/0	
TpsbE-1075	63997	-			*petL*		CTACATATGTTGGTAAAGCATTTACCTAAGTTATGGAAAATT*CATA*ATTCA	6/0	
TpetG-157	64202	+	*petG*				TGGCTTGAACAAGATACGTCTTATTTGAATTGACTGAATAATAATAATTCA	2/0	
TtrnW+1	64663	-	*trnW-CCA*				CGTGAGATACATTAGGAATAGAAAAAGATTTTCTTTTTGAATGAATGAAAG	8/1	TSS at +1 of *trn* gene
TtrnP-21	64898	-	*trnP-UGG*				AAGAAAATTATGATGTGGAAAAGAAGACAGGAATTGTGTACAATGGCATTG	1/0	
TpsaJ-251	64966	+	*psaJ*		*trnP-UGG*		TTCCTGTCTTCTTTTCCACATCATAATTTTCTTATTCTTTCTC*TATC*TATA	42/4	
Trps18-189	66073	+	*rps18*				AACAGAAAGAGTAAGAACTTCATATTTTATATATAAATAAAA*CATA*GTATA	7/2	disconnected
TpsbB-3143	66268	+		*rps18*			AAAGAAGTGCATTTTCCAATTTTAAATTAAATAAGGAATAAAT*CATG*TATA	9/1	
TtrnP-1937	66814	-			*rps18*		TAAATTAAGATTAAGTGGTAGGAATCGACGAGCTGGATTACTTTCTTTATA	65/8	
Trpl20 -91	67387	-	*rpl20*				CGTGGTTGGGAAGGTTACTATAGCAAAAGCCATTGGAATTAA*TATT*TTATA	3/1	
TpsbB-1766	67645	+				x	TTACGAACTTGTATAGTTCACATTCTTTGGCTCTACCTATCCAT*TATA*GAG	6/0	minor 5' ends: 67645, 67647

Trpl21-376	67672	-	*rpl20*				GGTCAGCTACTTAGCCAACTTTCATAATTAAATACCGTCACTG*TATG*GATA	3/0	disconnected
Trps12-5	68113	-	*rps12*				ATCATCAGGTTAAGATGGATCTAAACCAATCCATTTTTTTCTA*TATA*CATA	8/2	
TclpP-132	69032	-	*clpP*				ATAACCTTTCCATCTATGTATTAATAGAATCTATAGTATTCA*TATA*GAATA	56/4	Hübschmann and Börner, 1998
TpsbB-376	69035	+	*psbB*				TATTGTCTTAATTTTTATTCTTATTCTATTCTATTTTATTCT*TATT*CTATT	3/0	disconnected
TpsbB-360	69051	+	*psbB*				ATTCTTATTCTATTCTATTTTATTCTTATTCTATTCTATATGAATACTATA	3/1	disconnected
TpsbB-355	69056	+	*psbB*				TATTCTATTCTATTTTATTCTTATTCTATTCTATATGAATAC*TATA*GATTC	8/0	disconnected
TclpP-182	69082	-	*clpP*				ATCACCCATTCCTTTTTCTTTATTCAATCTGTCTTACCTTTCT*TATA*TGTA	3/0	
TpsbB-324	69087	+	*psbB*				TATATGAATACTATAGATTCTATTAATACATAGATGGAAAGGT*TATA*CATA	4/1	disconnected
TclpP-1184	70084	-			*psbB*		TACTACTGGAAAGTACAGTTTCAATATTGCCCATACGTAATCCTT*TATA*TA	2/0	
TpsbT-17	71075	+	*psbT*				AAATTCTCCCAAATGACAAATGAATAGGTGTGGAAGTTATAAT*TGTA*AATA	6/0	
TpsbN-46	71434	-	*psbN*		*psbH ; psbT*		ATTTAATCATTGGTGTTGACTTTGTATACTATTCCGTTGTAGT*TGTA*AATA	2/0	
TpsbH-28	71464	+	*psbH*		*psbN*		GGATCTACGAAAAGATCGTGTATTTACAACTACAACGGAATAG*TATA*CAAA	2/0	
TpsbN-92	71480	-	*psbN*		*psbH*		CCTAGGTTTAGAACTATCTTCAACGGTTTGTGTAGCCATAAAT*TATA*TTTA	2/0	disconnected
TpsbN-488	71876	-	*psbN*		*petB*		ACGAATTAGAATCATCTTGATAGAGATAACAGGCCCTTTCTA*TGTA*TTATC	2/0	disconnected
TpetD-95	73335	+	*petD*				TCGGGTAGGTTGTGGTATTTCATTGCTACAAGCATGGGTTAT*TGTA*AAATA	6/3	
TtrnH-7906	73689	+		*petD*			AGAGTGGAAATGGAATCAGTCATGTGACATGATCCAATTCTC*TATT*TATTA	4/0	
TtrnH-7547	74048	+		*petD*			GTGCAACTTTGGAAAAGAGAATCTGATAAAGCTTTTTTTGTCTAGAGTCAT	3/0	
TpsbN-3371	74759	-			*petD*		TTGCATCTTTGGTACAATCTATATTTTCGCGAAATGGATCATAATAAAATA	93/9	
TpsbN-3547	74935	-		*rpoA*			ATCTTATGAGAATTGACAGTTTTCGTATGGAAGATAGAAAACA*TATA*TGGG	3/0	
TpsbN-3833	75221	-		*rpoA*			ACCCCTAAGGAAGCGCTTTATGAGGCTTCTCGTAATTTGATTGATTTATTT	2/0	
TtrnH-6209	75386	+			*rpoA*		GAAAAAATACTATAATTCACATTTCGAACAGGCATGGATACAG*CATC*TATA	13/1	
TrpoA-57	75955	-	*rpoA*				AATGGATGTCGACCACCTAAAAAAAGACGTCTGTAAAAGAATTAACCCGCT	3/0	disconnected
TrpoA-340	76238	-	*rpoA*	*rps11*			TCATGTTCAAGCTAGTTTCAACAATACCATTATAACTGTTACAGACCCACA	4/0	
TinfA-84	77227	-	*infA*	*rps8*			TCGAGAAGCTCGACTAAACAGAATTGGGGGAGAAGTCTTATGT*TATA*TATG	3/0	
Trps8-142	77775	-	*rps8*				GAATTGAATTTTACCAAAATAGTTTCATTAGCTCCTGAAGTAT*TATA*AATA	16/5	disconnected
Trps8-321	77954	-	*rps8*	*rpl14*			CCCAAATGCCTCTAGAAAGATCCGAAGTAATTCGAGCTGTAAT*TGTA*CGTA	4/1	disconnected
Trpl14-883	79032	-		*rpl16*			AAATAAGACAGGAAAGAGTAAATATTCGCCCGCGAAATTCT*TATT*GAATTA	8/0	
Trpl14-933	79082	-		*rpl16*			GGCGGAACAAACCAAAAAAATTGTCTTATTTGGTAAGTTTATAAATTAACA	2/0	

TtrnH-2014	79581	+			*rpl16*		AGATAAAAAAGTGCAGTTGCTACAACTATATGACACATCTACT*CATT*TATG	2/0	
Trpl14-1440	79589	-		*rpl16*			CTTCGTATTGTCGAGATCCTAACTAAGAAACAGTCACTATATGAATAAATA	3/0	
TtrnH-1990	79605	+			*rpl16*		ACTATATGACACATCTACTCATTTATGATAGAGGTATTTATT*CATA*TAGTG	2/0	
Trpl14-1499	79648	-		*rpl16*			TTAAAGAAAAAAACTTAGTAATTATAGAAAATTAACTAATACTAATAACCA	6/1	
Trps19-75	81538	-	*rps19*		*trnH-GUG*		CCAGCTAAAGGATTTTTTTCTTTTTTCCGTTGATCATTATTC*TATT*TATTC	11/1	minor 5' ends: 81538, 81539
TtrnH+1	81595	+	*trnH-GUG*				TAGAATAATGATCAACGGAAAAAAGAAAAAAATCCTTTAGCTG*GATA*AGGG	345/29	TSS at +1 of *trn* gene
Trps19-148	81611	-	*rps19*		*trnH-GUG*		ACGGGAATTGAACCCGCGCATGGTGGATTCACAATCCACTGCCTTGATCCA	2/0	disconnected
Trps19-208	81671	-	*rps19*		*trnH-GUG*		CGTAACTAGGAATATGGAAAATTGCATTTTTGGAATTTGCAATAATGCGAT	2/0	disconnected
TtrnV-10461	81923	+			*rpl2*		GGGTTCATAACTACCCCTCTTACTACGGGGCGTTTACCTAGCCAACACTTA	3/0	
Trps19-802	82265	-		*rpl2*			CCTTTCCTATATATCCACTCATGTGGTACTTCATCATATGATT*CATA*TAAG	8/0	
TtrnV-9816	82568	+				x	TTTGCCTCTGACCATCAAATTAAATGTGAATAACCCGTCCTCCTCTCTTTG	2/0	
Trps19-1464	82927	-		*rpl2*			CCATAGAATACGACCCTAATCGAAATGCATACATTTGTCTCATACACTATG	3/0	
Trps19-1515	82978	-		*rpl2*			GTAAAATAGATTTTCGACGGAATCAAAAAGACATATCTGGTAGAATCGTAA	2/0	
Trpl2+1	83209	-	*rpl2*				CTATTCCACTTCTAGATAGAGAAAAAAACTAAAGGAGAATACTTAATAATA	2/0	
Trpl2-33	83242	-	*rpl2*	*rpl23*			GACGTATGATCATTACCCTTCAACCGGGTTATTCTATTCCACTTCTAGATA	3/0	
Trpl23-71	83580	-	*rpl23*				TCTCATCCATCATCCATACATAACGAATTGGTATGGTATATT*CATA*CCATA	523/53	Hübschmann and Börner, 1998
TtrnV-7711	84673	+				x	TTGTACAAAATAGACTTACTAAGTAATAACCCCATAGAATCTATC*TATA*TA	4/0	
TtrnI-1701	85458	-				x	GCCGACCAAAAATTCTATATGTCTATTCTATCTATGATATTTC*TATA*TATA	13/1	
TtrnV-6904	85480	+				x	CGGATGTCATATTAGATATCATATTCTATATATATAGAAATAT*CATA*GATA	3/1	
TtrnL+1	86217	-	*trnL-CAA*				ATCACATTAGATAGATATCATATTCATGGAATACAATTCACTTTCAAGATG	1053/58	TSS at +1 of *trn* gene
TtrnL-48	86265	-	*trnL-CAA*				GATCGATCGAAATACTCCAAGACTCCACCTTTGTCATATATTC*CATA*TATC	1/0	TSS of *trn* precursor
TtrnL-56	86273	-	*trnL-CAA*				ATATGACCGATCGATCGAAATACTCCAAGACTCCACCTTTGT*CATA*TATTC	1/0	
TtrnV-5120	87264	+			*ndhB*		ACAATGCAAGCAAAAGTTCCTAGATTCATGGAGATATAGAACAG*CATA*TAA	2/0	
TtrnL-1101	87318	-		*ndhB*			AGCATGAAACGTATGCTTGCATATTCGTCCATAGGGCAAATCG*GATA*TGTA	2/0	
TtrnL-1767	87984	-		*ndhB*			CGTTATTACGGGTAGTTCCTACAAAGAATCGGACTAATGACG*TATA*CAATG	3/0	

TtrnL-1839	88056	-		*ndhB*			AATAACAATTAAGGTGAAGCAGGGTCAGGAACAACGAATCTCTTTATGATA	3/0	
TndhB-95	89129	-	*ndhB*				GGGCTTGCTTAAGAATAAGAAAGAAGAATCTTATGGAAATAG*CATG*GAATA	12/2	
TndhB-275	89309	-	*ndhB*				ATAGAGCTCTTGCACATTTTCGTTAATCCATGAACAGAATCTA*TGTA*TGTA	37/16	disconected
Trps12-204	90862	-	*rps12*				GATCCGCATATGTTTGGTAAAAGAACAATCTTCTCCTTTAAT*CATA*TCATA	2/0	
TtrnV-1514	90870	+				x	AATTGAGTCACGTTTTCATGTTCTAATTGAACACTTTCCATT*TATG*ATATG	8/0	potential asRNA torps12 5' UTR?
Trps12-266	90924	-	*rps12*				AAGAATGAGGGGCAAGGGGATTGATACCGAGAAAGATTTCTTCTTATTATA	6/0	disconnected; minor 5' ends: 90923, 90924
TtrnV-1332	91052	+				x	TTCCATTTCTTCATTTGGAATCTGGGCTCTTCTATCTTCGACT*TATT*TTTT	2/0	
TtrnV-1019	91365	+				x	CCGAGATCTTTTCGATGACCTATGTTGTGTTGAAGGGATATC*TATA*TGATC	2/0	
Trps12-747	91405	-	*rps12*				TAATATATAGATAATCGAAATTGAAAAGAACTGTCTTTTCTG*TATA*CTTTC	6/0	disconnected
TtrnV-929	91455	+				x	GAAAGTATACAGAAAAGACAGTTCTTTTCAATTTCGATTATC*TATA*TATTA	29/1	
Trps12-827	91485	-				x	GCCGGGATCGGTAACTAATAGAATAGTACTACTAACTAATACTAA*TATA*TA	51/11	potential distant TSSof *rps12*?
Trps12-853	91511	-				x	TCATCACGGAAGAAAGAACTCACAGAGCCGGGATCGGTAACT*AATA*GAATA	3/0	potential distant TSS of *rps12*?
TtrnV-30	92354	+	*trnV-GAC*				AGCCCGGAGGAAGAGTGGCCTTGAGTTTCTCGCCCCTTTGCCTTAGGATTC	2/0	disconnected
TtrnV+1	92384	+	*trnV-GAC*				CGCCCCTTTGCCTTAGGATTCGTTAATTCTCTTTCTCGATGGGACGGGGAA	29/3	TSS at +1 of *trn* gene
Trps12-2002	92660	-			*rrn16*		AGCCAGGATCAAACTCTCCATGAGATTCATAGTTGCATTACT*TATA*GCTTC	2/0	
Trrn16-13	92672	+	*rrn16*				GGAAAGACAATTCCGAATCTGCTTTGTCTACGAATAAGGAAGC*TATA*AGTA	6/2	
Trps12-8997	99655	-				x	TCTTGACTGAAAGGGACACCAAAGGCCTCTGCCCTCCCTCTC*TATC*TATCC	4/0	minor 5' ends: 99655, 99656
Trps12-9058	99716	-			*rrn5*		TCGAACCATGAACGAGGAAAGGCATGAGATAAATATTGGCTAGTAATTGTG	2/0	
Trps15-1328	100754	+				x	AGCACTTCCTATCATTTAATATCCATCCCTTTGGTCTTATTGA*CATA*AGAG	3/0	
Trps15-1293	100789	+				x	CTTATTGACATAAGAGATGTCATTTATAGTCTATCTCTTTCTA*TATA*TGGA	6/2	
Trps15-921	101161	+				x	TCTTAGCTAAACAGGTGGAAGATCTATCCAATTTGGTTATAT*TATA*TCATG	11/0	
TtrnN-775	101309	-				x	CTTCGACTTCTATTAGTTTCTTTTCTTCTTTAATGCAATAGC*TATA*GTTTG	2/0	
Trps15-772	101310	+				x	TTCCTATAAGAGAATGGTTTCCATTACTTTGAGAAATGGATTCT*TATA*TCA	5/0	
Trps15-751	101331	+				x	CATTACTTTGAGAAATGGATTCTTATATCAAACTATAGCTATTG*CATT*AAA	3/1	
Trps15-738	101344	+	*rps15*				AATGGATTCTTATATCAAACTATAGCTATTGCATTAAAGAAGAAAAGAAAC	3/0	disconnected;

Trps15-545	101537	+	*rps15*				TTGGAAAGTTATGGAAGGAGACCCATCATTTTGCAATGAAAACAA*CATA*TA	12/6	disconnected; minor 5' ends: 101537, 101538
Trps15-237	101845	+	*rps15*				TTCATTTTGCTAAATGATATCAATTAAATGGTGTATCAATTC*CATA*AATTG	5/0	disconnected
Trps15-228	101854	+	*rps15*				CTAAATGATATCAATTAAATGGTGTATCAATTCCATAAATTG*CATA*TAGCA	485/35	disconnected; minor 5' ends: 101851, 101852, 101854
TtrnN-1479	102013	-			*rps15*		AAAATAAAAAATTATGTCAGTTATTTTGAAGTTATTCTAATCT*CGTA*CACA	70/14	
Trps15-40	102042	+	*rps15*				ATGAATAATTGCAAATTTGTGTGTGTACGAGATTAGAATAACTTCAAAATA	2/0	
TtrnN-1774	102308	-			*rps15*		TAAATTACTGCTCCCGAATATTCAACTGACCGATTAATTTCT*TATA*ACGTA	5/0	
Trpl32-1829	104014	+			*ndhF*		GAATAAAGCAGCTTGATAAGAACCTATACCTAGAGCTAACAT*CATA*TAACC	29/3	minor 5' ends: 104013, 104014
Trpl32-1224	104619	+			*ndhF*		TGGAGCTAGTAACCAATCCCAACATAGAAGTATTGAAAAAACT*TATA*TAAA	15/3	
Trpl32-990	104853	+			*ndhF*		CATTTATTTGTTGAATAGACAGTTGAACTGAGAATACCATAGC*TATA*CTTA	3/0	
Trpl32-267	105576	+	*rpl32*				TTCTTTTTATCCCCCCGGGGGATAGGCTTCCATACCATATATC*TATA*TATG	8/0	disconnected; minor 5' ends: 105576, 105578
TndhF-596	105591	-	*ndhF*				TTTTAATAATATAGAATATATAGAAGGGTTTCCCATTTATATC*TATT*TATA	4/0	disconnected
Trpl32-249	105594	+	*rpl32*				GGGATAGGCTTCCATACCATATATCTATATATGGAGTATACTTGA*TATA*TA	4/0	disconnected
Trpl32-84	105759	+	*rpl32*				AGTAAAAAATAATTCACAATTTCAGTATCTTTCAGTATCTAAG*TATA*AATA	2/0	
TtrnL-586	106161	+	*trnL-UAG*				TGTCGAATTACTTCGGTACAATATTCTTATAACAAACCCCTCT*TATA*TATA	148/19	potential ncRNA candidate?
TtrnL-1	106746	+	*trnL-UAG*				TTCCCGAGATAATAGCTATTATTCTTTTAAGTTACCTATTATTTGAAGTTA	4/0	
TndhF-2814	107809	-			*ccsA*		GGTGTTAATGTAAACGAACCATAACTATGTAAACCTATTCCT*AATA*GATTG	2/0	
TtrnN-8049	109551	+			*ndhD ; psaC*		ACAACAAGTATTGTTAACCAAGGAAAATAACTCATGATAAAGTGATAAAGA	20/0	
TndhD-331	109866	-	*ndhD*	*psaC*			TATAGATCCAATGTCACATTCCGTAAAAATTTATGATACATG*TATA*GGATG	18/0	disconnected
TtrnN-7020	110580	+			*ndhE*		GACTATCAAATAGATCAGAAAATGTTACGAGATTTAGATTAATTGAATTCA	2/0	
TndhG-111	111547	-	*ndhG*				TTTTCTTATTTTAAATCAGACTAGATTTTGATGATATAATATAGTTTATAA	8/2	minor 5' ends: 111547, 111548
TtrnN-5275	112325	+			*ndhA*		TGGGAGAGCTTTTGTGTTGAAAATATTCTTACTATCTATTAT*TGTA*TTATC	3/0	
TndhI-99	112327	-	*ndhI*				CTGGGCAATCTATTATTAACAACCTCTTCTCAACTTGTTTCAC*TATA*AATA	56/7	

TndhI-217	112445	-	*ndhI*	*ndhA*			TACATTAACTAAAGCTTATTTATTTCTCTTCATTTCTATCACAATAAGATG	3/0	disconnected
TtrnN-4707	112893	+			*ndhA*		ATCAACTGTACTTGAACTGTTAGATAATCATAGTCGATGATAA*CATC*ACAG	2/0	
TndhI-875	113103	-		*ndhA*			CAAGATTAAGTTATTACTGAACAAAGAAAAATTTGGAATATAT*TATA*AAGG	2/0	
TtrnN-2842	114758	+			*ndhH*		TTATACTCAAAATCATTCCATTCAGAATTCTTTTCTTTCTTAAAGCGTCGG	2/0	
TndhA-980	115430	-		*ndhH*			GAAAAAATCGCGGAAAACCGAACGATTATACAATACTTACCT*TATG*TAACA	2/0	

Appendix B: Verification of TSSs by 5'-RACE analysis. TSSs (with various abundance in dRNA-seq libraries) belonging to different categories were selected for verification by 5'-RACE. The name, type, genomic location, abundance and library of the analyzed TSSs are given. The genomic positions of additional TSSs mapped by 5'-RACE analysis are also included. The number of clones supporting a given TSS in T+ (TAP+) and T- (TAP-) reactions are listed. The TSSs verified successfully as major 5'-PPP ends are marked in bold. The primary transcripts with a stable structure formed immediately downstream their TSSs are underlined.

		Name	TSS type	Genomic location based on dRNA-seq	Strand	No of cDNAs (+/-)	Library	5'-RACE Confirmation of dRNA-seq TSS		Location of additional TSS based on 5'-RACE (sample)	5'-RACE confirmation	
								T+	T-		T+	T-
aTSSs and oTSSs	1	**T*trnG*-1652**	asTSS_*trnE*	15733	-	13/3	W	31/31	22/32			
	2	T*trnG*-1659		15740		8/1	G	5/30	0/30	15733 (G)	6/30	1/30
	3	**T*trnC*-12426**	asTSS_*atpI*	31347	-	11/2	G	16/28	12/31	31346 (G)	8/28	4/31
										31342 (W)	18/32	1/32
	4	**T*trnT*-2150**	oTSS	48681	-	15/2	W	13/22	2/28			
	5	**<u>T*trnM*-582</u>**	asTSS_*trnV*	51481	+	51/11	G	27/30	0/30	51411 (G)	10/29	0/31
	6	**T*psbN*-495**	asTSS_*petB*	71883	-	41/9	G	16/31	0/30	71870 (G)	9/31	0/30
										71875 (W)	4/30	1/31
										71876 (W)	5/30	1/31
										71877(W)	5/30	2/31
	7	**T*trnH*-6209**	asTSS_*rpoA*	75386	+	13/1	W	14/30	0/32			
	8	T*trnI*-1694	oTSS	85451	-	4/0	G	4/32	0/29	85439 (G)	8/32	0/29
	9	**T*trnI*-1701**		85458		13/0	W	14/27	9/30	85451 (W)	4/27	1/30
										85456 (W)	4/27	1/30
	10	**<u>T*trnV*-1019</u>**	oTSS	91365	+	2/0	W	7/7	-			

aTSSs and oTSSs	11	**T*trnN*-1479**	asTSS_*rps15*	102013	-	70/14	W	25/25	2/31	101970 (G)	3/30	0/29
										101982 (G)	3/30	0/29
										101986 (G)	5/30	1/29
										102013 (G)	5/30	0/29
	12	T*rpl32*-1830	asTSS_*ndhF*	104013	+	2/0	G	1/14	-	104011 (G)	7/30	-
										104012 (G)	4/14	-
										104040 (G)	6/30	-
	13	**T*rpl32*-1829**		104014		29/3	W	20/31	-	104013 (W)	3/31	-
	14	**T*rpl32*-1224**	asTSS_*ndhF*	104619	+	15/3	W	22/31	20/31	104618 (W)	4/31	2/31
										104619 (G)	10/31	-
										104618 (G)	5/31	-
	15	**T*trnN*-5275**	asTSS_*ndhA*	112325	+	3/0	W	5/5	-			
gTSSs	16	**T*psbA*-80**	gTSS_*psbA*	1760	-	7938/1235	G	16/16	-			
	17					141/9	W	16/16	-			
	18	T*trnE*+1	gTSS_*trnE*	15791	+	2729/529	G	6/10	7/9			
	19					444/17	W	8/10	10/10			
	20	**T*atpH*-212**	gTSS_*atpH*	32447	-	4/0	W	4/7	-			
	21	**T*psaA*-209**	gTSS_*psaA*	42089	-	263/369	G	2/2	-			
	22	**T*ndhC*-249**	gTSS_*ndhC*	50708	-	7/1	W	8/8	-			
	23	**T*ndhC*-329**		50788	-	2/0	W	2/7	0/8			
	24	**T*ndhC*-336**		50795	-	4/0	W	5/7	0/8			
	25	T*rbcL*-426	gTSS_*rbcL*	54513	+	3/0	G			54622 (G)	25/31	0/12
	26	**T*rbcL*-214**		54725		7/2	W	17/31	1/31	54754 (W)	10/31	2/31
	27	T*psaI*-83	gTSS_*psaI*	57899	+	5/0	W	2/31	0/31			
	28	T*psaI*-74		57908		3/0	W	3/31	0/31	57919 (W)	15/31	6/31
	29	**T*psaI*-70**		57912		49/42	G	17/31	-	57913 (G)	6/31	-
										57914 (G)	5/31	-
	30	**T*petL*-93**	gTSS_petL	63991	+	1/12	G	3/13	0/14			
	31	**T*petG*-39**	gTSS_*petG*	64320	+	6/7	G	4/8	2/8	64318 (G)	2/8	1/8

gTSSs	32	**T*psaJ*-251**	gTSS_*psaJ*	64966	+	42/4	W	8/16	3/15			
	33	**T*psaJ*-247**		64970		2/0	G	2/9	-			
	34	T*psaJ*-198		65019		5/2	G	1/7	-	65016 (G)	4/9	-
	35	T*psbB*-376	gTSS_*psbB*	69035	+	3/0	W	2/26	-			
	36	**T*psbB*-360**		69051		3/1	W	9/26	-			
	37	T*psbB*-355		69056		8/0	W	0/26	-			
	38	**T*psbB*-324**		69087		4/1	W	5/16	0/16			
	39	**T*psbB*-176**		69235		91/23	G	14/16	1/15			
	40	**T*psbN*-46**	gTSS_*psbN*	71434	-	19/329	G	14/16	2/14			
	41					2/0	W	14/30	1/29			
	42	**T*ndhB*-275**	gTSS_*ndhB*	89309	-	2/12	G	4/7	-			
	43					37/16	W	6/8	-	89308 (W)	2/8	-
	44	**T*ndhI*-79**	gTSS_*ndhI*	112307	-	41/32	G	7/8	-			
	45	T*ndhI*-99		112327		4/10	G	0/8				
	46	**T*ndhI*-99**				56/7	W	6/8	-			
iTSSs	47	**T*rpoC1*-979**	iTSS_*rpoB*	22376	+	6/1	W	2/4	-			
	48	**T*rpoC1*-599**	iTSS_*rpoB*; gTSS_*rpoC1*	22756	+	23/3	W	16/16	-	22756 (G)	11/16	-
										22758 (G)	4/16	-
	49	**T*ndhK*-158**	iTSS_*ndhC*; gTSS_*ndhK*	50264	-	9/4	G	2/7	0/6	50280 (G)	3/7	0/6
	50	**T*petL*-3023**	iTSS_*petA*	61061	+	63/4	W	14/15	-			
	51	**T*trnL*-838**	iTSS_*ndhB*	87055	-	11/0	G	9/14	-	87057 (G)	4/14	-
	52	**T*ndhD*-331**	iTSS_*psaC*; gTSS_*ndhD*	109866	-	18/0	W	15/16	-			

Appendix C: Identical TSSs in G and W dRNA-seq libraries. The name, genomic location, strand, number of cDNAs in (+) and (-) libraries, and 40 nt upstream sequence of the 22 identical TSSs in G and W dRNA-seq libraries are given. The predicted PEP and NEP promoter elements are underlined and highlighted in the upstream sequence, respectively. The nature of each TSS is discussed in the *Comments* column. G= green plastids

Name	Genomic location	Strand	TSS type	No of cDNAs (G+/G-)	No of cDNAs (W+/W-)	Sequence -40 nt upstream + TSS (41nt)	Comments
T*psbA*-80	1760	-	gTSS_*psbA*	7938/1235	141/9	TGGTTGACATTGGTATATAGTCTATGTTATACTGTTAAATA	PEP transcript in G
T*trnK*-239	4707	-	gTSS_*trnK*	2/1	3/0	AATGATAAGGGTGTTCCTCTTGCATGTATTCTCATACAATA	Potential NEP transcript in G
T*psbK*-783	6484	+	oTSS	2/3	39/3	GTTTAATTCATTTAATTACTAGAATTAGAATTCTATTAGTA	Potential NEP transcript in G
T*trnS*+1	8177	-	gTSS_*trnS*	1434/178	213/14	TGCCTATATCATATCACGGAAACCTTTCGCTTTGGAACGTG	TSS at +1 relative to *trn* gene start
T*trnfM*+1	13239	-	gTSS_*trnfM*	6330/790	162/9	TATTCAAGCCTTTTTTGTCCACCAGTTTCTGGTACTACAGA	TSS at +1 relative to *trn* gene start
T*trnE*+1	15791	+	gTSS_*trnE*	2729/529	444/17	TAATCACGAGCGGTTGTATATGGCCCTATCGTCTAGTGATG	TSS at +1 relative to *trn* gene start
T*psbM*-348	16868	+	gTSS_*psbM*	0/7	11/4	CTATGTGACCCATAGAAAGTTGCTCATATAATACATACATA	Potential NEP transcript in G
T*rpoB*-147	19940	+	gTSS_*rpoB*	12/9	223/12	TCGAAATGGTCTCTATTCATATGTATGAAATACATATATGA	NEP transcript in G
T*rps2*-152	30221	+	gTSS_*rps2*	11/10	16/1	GTTAATTCATTAAATTAAGGTTTTGTTTATACCATGTATCA	Potential NEP transcript in G
T*psaA*-209	42089	-	gTSS_*psaA*	263/369	5/0	ATGTCCGTTAGGCACCTAACCTTTATGTCATAATAGATCCG	PEP transcript in G
T*ndhC*-336	50795	-	gTSS_*ndhC*	2/0	4/0	ATTCTCATTTTTATTTAATAGTCTCTTATTATTATTAAATA	Potential NEP transcript in G
T*trnP*-21	64898	-	gTSS_*trnP*	11/2	1/0	TGATGTGGAAAAGAAGACAGGAATTGTGTACAATGGCATTG	Potential NEP transcript in G
T*trnP*-1937	66814	-	aTSS_*rps18*	8/8	65/8	TTAAGTGGTAGGAATCGACGAGCTGGATTACTTTCTTTATA	Potential NEP transcript in G
T*psbN*-46	71434	-	gTSS_*psbN*; *aTSS_psbH*; *aTSS_psbT*	19/329	2/0	TGGTGTTGACTTTGTATACTATTCCGTTGTAGTTGTAAAATA	PEP transcript in G
T*psbN*-3371	74759	-	aTSS_*petD*	62/29	93/9	GGTACAATCTATATTTTCGCGAAATGGATCATAATAAAATA	Potential NEP transcript in G
T*rps8*-142	77775	-	gTSS_*rps8*	2/0	16/5	TTACCAAAATAGTTTCATTAGCTCCTGAAGTATTATAAATA	Potential NEP transcript in G
T*rpl23*-71	83580	-	gTSS_*rpl23*	2/1	523/53	CATCCATACATAACGAATTGGTATGGTATATTCATACCATA	NEP transcript in G
T*trnL*+1	86217	-	gTSS_*trnL*	5038/750	1053/58	ATAGATATCATATTCATGGAATACAATTCACTTTCAAGATG	TSS at +1 relative to *trn* gene start
T*ndhB*-275	89309	-	gTSS_*ndhB*	2/12	37/16	TGCACATTTTCGTTAATCCATGAACAGAATCTATGTATGTA	Potential NEP transcript in G

T*trnV*+1	92384	+	gTSS_*trnV*	28/9	29/3	CCTTAGGATTCGTTAATTCTCTTTCTCGATGGGACGGGGAA	TSS at +1 relative to *trn* gene start
T*rps15*-228	101854	+	gTSS_*rps15*	53/44	485/35	TCAATTAAATGGTGTATCAATTCCATAAATTG*CATA*TAGCA	NEP transcript in G
T*ndhI*-99	112327	-	gTSS_*ndhI*	4/10	56/7	TATTATTAACAACCTCTTCTCAACTTGTTTCAC*TATA*AATA	Potential NEP transcript in G

Appendix D: Operon map of the barley chloroplast genome. The operon organization of the barley plastome and all gTSS mapped by dRNA-seq are listed in the table below. Genes clustered in an operon are grouped together and highlighted in gray nuances. Genes transcribed as monocistronic transcripts have no colored background. In the cases when there is more than one TSS assigned to a certain gene, the most abundant one is given in bold. TSS found identical in both green and white libraries are underlined.

Gene	Green plastids	White plastids
psbK	T*psbK*-171	T*psbK*-169
psbI	-	T*psbI*-178
psbD	T*psbD*-716; T*psbD*-711; **T*psbD*-557**	**T*psbD*-216**; T*psbD*-68
psbC	T*psbC*-194	T*psbC*-93
psbZ	-	T*psbZ*-451
trnG-GGC	T*trnG*-90	T*trnG*-82
trnT-GGU	**T*trnT*-70**; P*trnT*-59	T*trnT*-78; T*trnT*-68
trnE-UUC	T*trnE*-46; **T*trnE*+1**	T*trnE*-412; T*trnE*-39; **T*trnE*+1**
trnY-GUA	-	-
trnD-GUC	T*trnD*-133	T*trnD*-34
psbM	T*psbM*-348; T*psbM*-162; **T*psbM*-111**	T*psbM*-356; **T*psbM*-348**; T*psbM*-338; T*psbM*-160
rpoB	**T*rpoB*-147**	T*rpoB*-345; T*rpoB*-275; T*rpoB*-270; T*rpoB*-161; T*rpoB*-156; **T*rpoB*-147**
rpoC1	-	T*rpoC1*-599
rpoC2	-	
rps2	T*rps2*-261; **T*rps2*-152**	T*rps2*-207; **T*rps2*-152**
atpI	T*atpI*-140; **T*atpI*-118**	T*atpI*-626; T*atpI*-123
atpH	T*atpH*-350; **T*atpH*-201**; T*atpH*-175	T*atpH*-228; T*atpH*-212
atpF	T*atpF*-388	-
atpA	-	-
trnS-GGA	T*trnS*-5	T*trnS*-522
trnL-UAA	T*trnL*-152	-
trnF-GAA	-	-
trnM-CAU	**T*trnM*-582**; T*trnM*-95; T*trnM*-32	-
rbcL	T*rbcL*-426	T*rbcL*-214
rpl23	T*rpl23*-75	**T*rpl23*-60**; T*rpl23*+1
psaI	T*psaI*-70	**T*psaI*-83**; T*psaI*-74
ycf4	T*ycf4*-301; **T*ycf4*-2**	T*ycf4*-125; T*ycf4*-3
cemA	-	T*cemA*-85; **T*cemA*-69**
petA	T*petA*-112	-

petL	T*petL*-93	T*petL*-444; T*petL*-108
petG	T*petG*-39	T*petG*-157
psaJ	T*psaJ*-247; **T*psaJ*-198**; T*psaJ*-38	T*psaJ*-251
rpl33	-	-
rps18	-	T*rps18*-189
psbB	T*psbB*-176	T*psbB*-376; T*psbB*-360; **T*psbB*-355;** T*psbB*-324
psbT	-	T*psbT*-17
psbH	-	T*psbH*-28
petB	-	-
petD	-	T*petD*-95
trnH-GUG	-	T*trnH*+1
trnV-GAC	T*trnV*-102; T*trnV*+1	T*trnV*-30; **T*trnV*+1**
rrn16	T*rrn16*-116	T*rrn16*-13
trnI-GAU	T*trnI*-608	-
trnA-UGC	-	-
rrn23	-	-
rrn4.5	-	-
rrn5	-	-
trnR-ACG	-	-
rps15	T*rps15*-393; **T*rps15*-228**	T*rps15*-738; T*rps15*-545; T*rps15*-237; **T*rps15*-228**; T*rps15*-40
ndhH	-	-
rpl32	-	**T*rpl32*-267**; T*rpl32*-249; T*rpl32*-84
trnL-UAG	**T*trnL*-608**; T*trnL*-23	**T*trnL*-586**; T*trnL*-1
ccsA	-	-
ndhH	-	-
ndhA	-	-
ndhI	T*ndhI*-99; **T*ndhI*-79**	T*ndhI*-217; **T*ndhI*-99**
ndhG	T*ndhG*-114	T*ndhG*-111
ndhE	-	-
psaC	T*psaC*-270	-
ndhD	-	T*ndhD*-331
ndhF	-	T*ndhF*-596
trnN-GUU	T*trnN*-225	-
rps12 3′	-	T*rps12* -747; T*rps12*-266; T*rps12* -204
rps7	T*rps7*-212	-
ndhB	T*ndhB*-275	**T*ndhB* -275**; T*ndhB* -95
trnL-CAA	T*trnL*-23; **T*trnL*+1**	T*trnL*-56; T*trnL*-48; **T*trnL*+1**
trnI-CAU	-	-
rpl23	T*rpl23*-71	T*rpl23*-71

rpl2	-	**T*rpl2*-33**; T*rpl2*+1
rps19	-	T*rps19*-208; T*rps19*-148; **T*rps19*-75**
rpl22	-	-
rps3	-	-
rpl16	T*rpl16*-98	-
rpl14	-	-
rps8	T*rps8*-281; T*rps8*-142	T*rps8*-321; T*rps8*-142
infA	-	T*infA*-84
rpl36	-	-
rps11	-	-
rpoA	-	T*rpoA*-57 (75955); **T*rpoA*-340**
psbN	T*psbN*-708; **T*psbN*-495**; T*psbN*-61; T*psbN*-46	T*psbN*-488; T*psbN*-92; T*psbN*-46
clpP	-	T*clpP*-182; **T*clpP*-132**
rps12 5′	-	T*rps12*-5
rpl20	T*rpl20*-334	T*rpl20*-376; P*rpl20*-91
trnP-UGG	T*trnP*-135; **T*trnP*-21**	T*trnP*-21
trnW-CCA	-	T*trnW*+1
psbE	T*psbE*-140	-
psbF	-	-
psbL	-	-
psbJ	-	-
atpB	T*atpB*-392; **T*atpB*-294**	**T*atpB*-593**; T*atpB*-450
atpE	-	-
trnV-UAC	T*trnV*-101; T*trnV*-89; **T*trnV*-4**	T*trnV*-111; **T*trnV*+1**
ndhC	T*ndhC*-376; T*ndhC*-336	T*ndhC*-639; T*ndhC*-336; T*ndhC*-329;**T*ndhC*-249**
ndhK	T*ndhK*-158	-
ndhJ	-	-
trnT-UGU	T*trnT*-9	-
rps4	T*rps4*-174	T*rps4*-178
ycf3	T*ycf3*-73	T*ycf3*-332; T*ycf3*-270; T*ycf3*-107; **T*ycf3*-82**
psaA	T*psaA*-209	T*psaA*-209
psaB	-	-
rps14	T*rps14*-733	-
trnfM-CAU	T*trnfM*-69	T*trnfM*-63
trnR-UCU	-	T*trnR*+1
trnC-GCA	T*trnC*-377	T*trnC*+1
petN	T*petN*-44	-
trnG-UCC	T*trnG*-3	T*trnG*-16
trnfM-CAU	T*trnfM*+1	T*trnfM*+1

trnS-UGA	T*trnS*-7; **T*trnS*+1**	-
trnS-GCU	T*trnS*-583; T*trnS*-485; T*trnS*-2; **T*trnS*+1**	T*trnS*+1
trnQ-UGG	**T*trnQ*-150**; T*trnQ*-21	-
rps16	-	T*rps16*-735; T*rps16*-438; **T*rps16*-206**; T*rps16*-95
trnK-UUU	**T*trnK*-337**; T*trnK*-239	T*trnK*-239
matK	T*matK*-295	T*matK*-260
psbA	T*psbA*-80	T*psbA*-146; **T*psbA*-80**

Appendix E: ncRNA candidates in barley mature chloroplasts. The name of the ncRNA candidates (denoted Hv_nc), name, type, strand, abundance in green and/or white libraries and genomic location of the corresponding TSS are given. Several candidates were selected for verification by 5'-RACE and 3' end mapping by 3'-RACE. The genomic location of 5' and 3' ends mapped by 5'/3'-RACE are included and the size of the analyzed ncRNA candidates is calculated. Multiple TSS were detected for several ncRNAs, with the most abundant one shown here in bold.

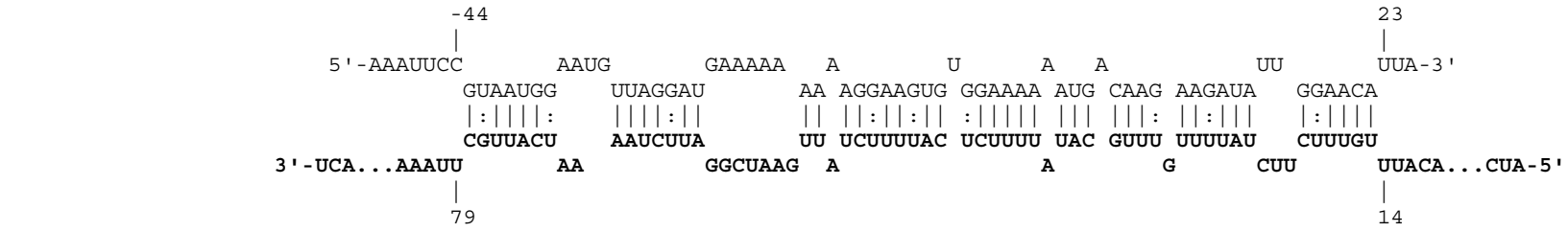
ncRNA name	TSS name	TSS type	Strand	No of cDNAs (G+/G-)	TSS based on dRNA-seq(G)	TSS based on dRNA-seq (W)	TSS based on 5'-RACE (G)	TSS based on 5'-RACE (W)	3' ends based on 3'-RACE (G)	size (nt)
Hv_nc1	T*psbK*-6580	as_*psbA*	+	3/0	687	-	-	-	-	-
Hv_nc2	T*psbK*-1093	as_*rps16*	+	3/0	6174	-	-	-	-	-
Hv_nc3	T*psbK*-783	oTSS	+	2/3	6484	6484	-	-	-	-
Hv_nc4	T*trnS*-1984	as_*psbC; as_psbD*	-	3/0	10161	-	-	-	-	-
Hv_nc5	T*trnT*-2278	oTSS	+	2/0	12925	-	-	-	-	-
Hv_nc6	T*trnT*-1457	as_*trnG-UCC*	+	2/0	13746	-	-	-	-	-
Hv_nc7	T*trnT*-1326	as_*trnG-UCC*	+	4/0	13877	-	-	-	-	-
Hv_nc8	T*trnT*-844	oTSS	+	3/0	14359	-	-	-	-	-
Hv_nc9	T*trnG*-1036	as_*trnT-GGU*	-	7/2	15117	15121	-	-	-	-
Hv_nc10	T*trnG*-1659	as_*trnE-UUC*	-	8/1	15740	15733	15740; 15733	15733	15597-15376	143-364
Hv_nc11	T*trnG*-1808	*as_trnE-UUC; as_trnY-GUA*	-	5/1	15889	-	-	-		
Hv_nc12	T*trnG*-2782	oTSS	-	4/0	16863	-	-	-	-	-
Hv_nc13	T*rpoB*-2539; T*rpoB*-2586	oTSS	+	47/10; 3/1	**17548**; 17501	17496	-	-	-	-
Hv_nc14	T*trnG*-3476; T*trnG*-3435	as_*psbM*	-	21/0; 7/8	**17557**; 17516	-	-	-	-	-
Hv_nc15	T*trnC*-808	oTSS	-	2/0	19729	-	-	-	-	-
Hv_nc16	T*trnC*-5056	as_*rpoC1*	-	3/3	23977	-	-	-	-	-

Hv_nc17	T*trnC*-9794	as_*rpoC2*	-	3/0	28715	-	-	-	-	-
Hv_nc18	T*trnC*-12426	as_*atpI*	-	11/2	31347	-	**31347**; 31346	31342	31203-30953	144-394
Hv_nc19	T*trnC*-13514	oTSS	-	3/0	32435	-	-	-	-	-
Hv_nc20	T*trnC*-14756	as_*atpF*	-	5/1	33677	-	-	-	-	-
Hv_nc21	T*trnC*-15956	as_*atpA*	-	2/0	34877	-	-	-	-	-
Hv_nc22	T*trnS*-7862	as_*psaB;as_rps14*	+	2/0	37295	-	-	-	-	-
Hv_nc23	T*trnS*-4772	as_*psaA*(1)	+	4/0	40385	-	-	-	-	-
Hv_nc24	T*trnS*-4375	as_*psaA*(2)	+	2/0	40782	-	-	-	-	-
Hv_nc25	T*trnS*-3158	as_*psaA*(3)	+	27/2	41999	-	-	-	-	-
Hv_nc26	T*trnS*-2690	as_*ycf3*(1)	+	2/0	42467	-	-	-	-	-
Hv_nc27	T*trnS*-814	*as_ycf3*(2)	+	2/0	44343	-	-	-	-	-
Hv_nc28	T*trnT*-1740	as_*trnF-GAA*(1)	-	12/3	48271	-	-	-	-	-
Hv_nc29	T*trnM*-3791; T*trnM*-3776	oTSS	+	10/5; 166	48272; **48287**	48282	-	-	-	-
Hv_nc30	T*trnT*-1812	as_*trnF-GAA*(2)	-	2/0	48343	-	-	-	-	-
Hv_nc31	T*trnM*-582	as_*trnV-UAC*	+	51/11	51481	-	**51481**; 51411	-	-	-
Hv_nc32	T*atpB*-857	as_*rbcL*(1)	-	2/0	55013	-	-	-	-	-
Hv_nc33	T*atpB*-1656	as_*rbcL*(2)	-	3/0	55812	-	-	-	-	-
Hv_nc34	T*atpB*-2409; T*atpB*-2468	as_*rbcL*(3)	-	29/3; 5/3	**56565**; 56624	56571	-	-	-	-
Hv_nc35	T*atpB*-5129; T*atpB*-5115	oTSS	-	3/0; 2/0	**59285**; 59271	-	-	-	-	-
Hv_nc36	T*petL*-2041	as_*psaJ*	+	7/8	62043	62051	-	-	-	-
Hv_nc37	T*petL*-1582	as_*psbF; as_psbL*	+	10/4	62502	-	-	-	-	-
Hv_nc38	T*psbE*-1628; T*psbE*-1633	as_*petG*	-	18/11 7/2	**64550**; 64555	-	-	-	-	-
Hv_nc39	T*trnP*-1928; T*trnP*-1937	*as_rps18*	-	48/46; 8/8	**66805**; 66814	66814	-	-	-	-
Hv_nc40	T*psbN*-495	as_*petB*(1)	-	41/9	71883	-	**71883**; 71870	**71877**; 71876; 71875	71835-71816	48-67

Hv_nc41	T*psbN*-708	as_*petB*(2)	-	4/0	72096	-	-	-	-	-
Hv_nc42	T*psbN*-908	as_*petB*(3)	-	3/0	72296	-	-	-	-	-
Hv_nc43	T*psbN*-1326	as_*petB*(4)	-	4/1	72714	-	-	-	-	-
Hv_nc44	T*psbN*-2112; T*psbN*-2132	as_*petD*(1)	-	3/0; 2/0	**73500**; 73520	-	-	-	-	-
Hv_nc45	T*psbN*-2355	as_*petD*(2)	-	3/0	73743	-	-	-	-	-
Hv_nc46	T*psbN*-3371	as_*petD*(3)	-	62/29	74759	-	-	-	-	-
Hv_nc47	T*trnH*-2693	oTSS	+	4/0	78902	-	-	-	-	-
Hv_nc48	T*trnV*-9666	as_*rpl2*	+	2/0	82718	-	-	-	-	-
Hv_nc49	T*trnI*-1090	oTSS	-	2/0	84847	-	-	-	-	-
Hv_nc50	T*trnI*-1694	oTSS	-	4/0	85451	85458	**85439**; 85451	**85458**; 85456; 85451	85251-85241	200-210
Hv_nc51	T*trnV*-6424	as_*trnL-CAA*	+	4/0	85960	-	-	-	-	-
Hv_nc52	T*trnV*-5213	as_*ndhB*	+	3/0	87171	-	-	-	-	-
Hv_nc53	T*rps12*-4898	as_*trnA-UGC*	-	3/0	95556	-	-	-	-	-
Hv_nc54	T*rps12*-9264	as_*rrn5*	-	7/0	99922	-	-	-	-	-
Hv_nc55	T*rps12*-9771	as_*trnR-ACG*	-	3/0	100429	-	-	-	-	-
Hv_nc56	T*trnN*-1479	as_*rps15*	-	0/0	-	102013	**102013**; 101986; 101982; 101970	102013	101885-101870	128-143
Hv_nc57	T*rpl32*-1830	as_*ndhF*(1)	+	2/0	104013	104014	**104011**; 104040; 104012	**104014**; 104013	104162-104209	151-198
Hv_nc58	T*rpl32*-1224	as_*ndhF*(2)	+	0/0	-	104619	**104619**; 104619	**104619**; 104619	104684-104701	65-82
Hv_nc59	T*trnN*-9189	as_*ndhD*	+	2/0	108411	-	-	-	-	-
Hv_nc60	T*trnN*-8056; T*trnN*-8047	as_*ndhD*; as_*psaC*	+	2/0; 9/4	109544; 109553	109551	-	-	-	-

Appendix F: Top 10 highest scoring candidate targets of potential *trans*-encoded ncRNA. IntaRNA, part of the Freiburg RNA tools (http://rna.informatik.uni-freiburg.de:8080/IntaRNAHelp.jsp; Smith, et al., 2010) was used to screen with default parameters the region from -50 to +25 relative to the start codon of annotated genes in the barely genome (NC_008590) for potential targets of the identified ncRNAs from intergenic regions in this analysis. The top 10 highest scoring candidates are listed. The energy score of the predicted interactions (underlined) is the sum of the hybridization free energy of the interacting sequences and the free energies required to unfold the interaction sites in both RNA molecules. The position of the interacting regions and the seed regions (regions of perfect complementarity) in both RNA molecules is given. The sequences of the ncRNAs are marked in bold.

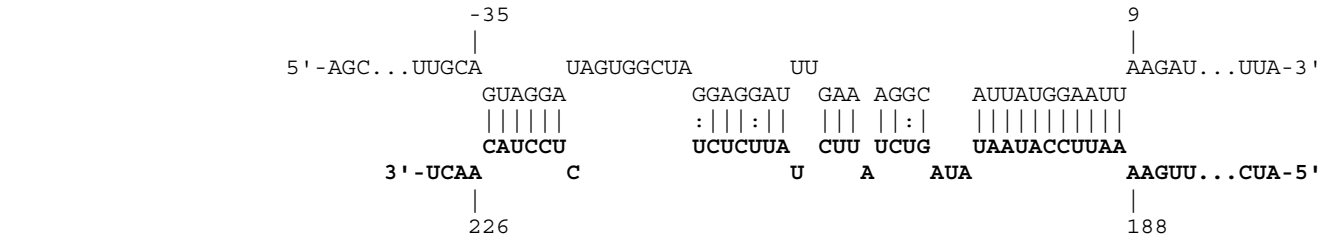
1. Predicted interaction between Hv_nc19 and *rps2*|Description:ribosomal_protein_S2



Energy:	-18.3 kcal/mol	Position - mRNA:	-43 -- 22
Hybridization Energy:	-35.2 kcal/mol	Position - ncRNA:	15 -- 78
Unfolding Energy - mRNA:	7.1 kcal/mol	Position Seed - mRNA:	-15 -- -9
Unfolding Energy - ncRNA:	9.9 kcal/mol	Position Seed - ncRNA:	45 -- 51

---

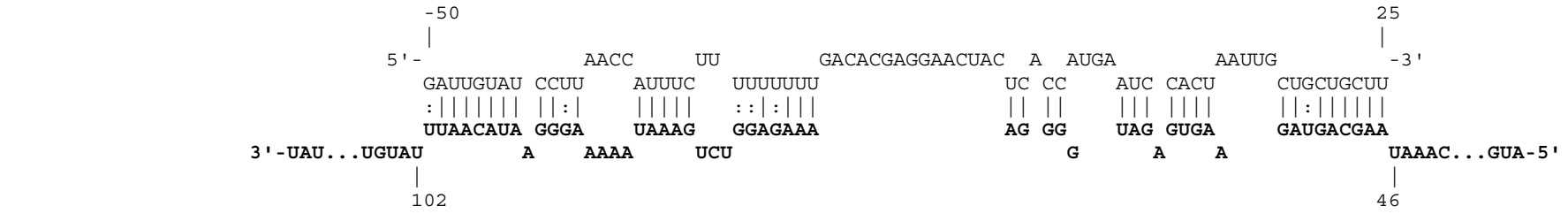
2. Predicted interaction between Hv_nc19 and *psaB*|Description:photosystem_I_P700_chlorophyll_ a_apoprotein_A2



Energy:	-17.3 kcal/mol	Position - mRNA:	-34 -- 8
Hybridization Energy:	-29.1 kcal/mol	Position - ncRNA:	189 -- 225
Unfolding Energy - mRNA:	8.8 kcal/mol	Position Seed - mRNA:	2 -- 8
Unfolding Energy - ncRNA:	3.0 kcal/mol	Position Seed - ncRNA:	189 -- 195

---

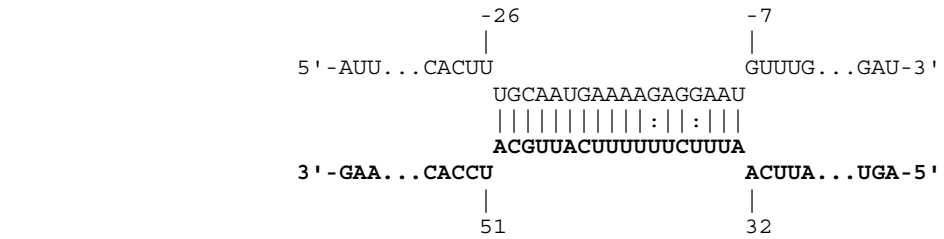
3. Predicted interaction between Hv_nc12 and *atpH*|Description:ATP_synthase_CF0_C_subunit



Energy:	-17.1 kcal/mol	Position - mRNA:	-50 -- 25
Hybridization Energy:	-33.4 kcal/mol	Position - ncRNA:	47 -- 101
Unfolding Energy - mRNA:	6.6 kcal/mol	Position Seed - mRNA:	19 -- 25
Unfolding Energy - ncRNA:	9.6 kcal/mol	Position Seed - ncRNA:	47 -- 53

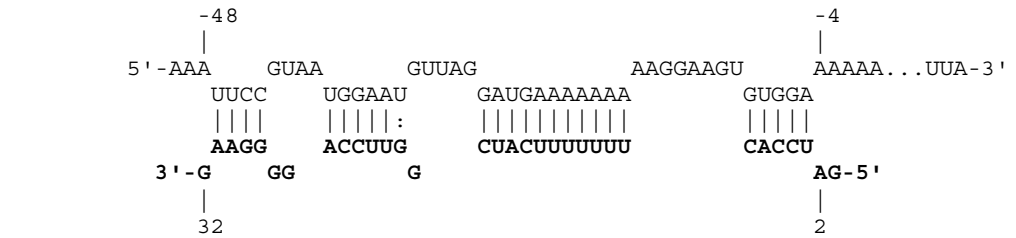
---

4. Predicted interaction between Hv_nc13 (T*rpoB*-2586) and rps16|Description:ribosomal_protein_S16



Energy:	-16.5 kcal/mol	Position - mRNA:	-25 -- -8
Hybridization Energy:	-22.3 kcal/mol	Position - ncRNA:	33 -- 50
Unfolding Energy - mRNA:	4.2 kcal/mol	Position Seed - mRNA:	-14 -- -8
Unfolding Energy - ncRNA:	1.6 kcal/mol	Position Seed - ncRNA:	33 -- 39

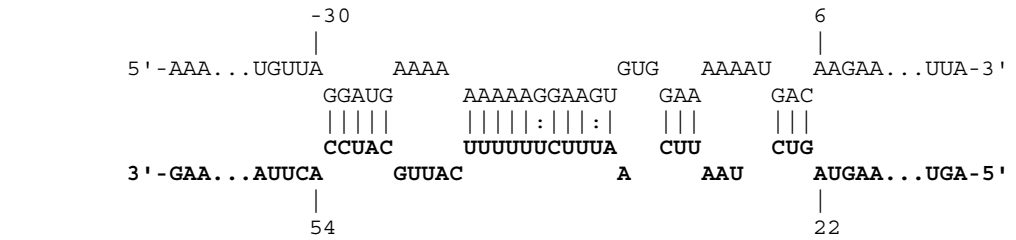
5. Predicted interaction between Hv_nc49 and *rps2*|Description:ribosomal_protein_S2|GeneID:4525123



Energy:	-15.7 kcal/mol	Position - mRNA:	-47 -- -5
Hybridization Energy:	-24.7 kcal/mol	Position - ncRNA:	3 -- 31
Unfolding Energy - mRNA:	5.6 kcal/mol	Position Seed - mRNA:	-24 -- -18
Unfolding Energy - ncRNA:	3.4 kcal/mol	Position Seed - ncRNA:	8 -- 14

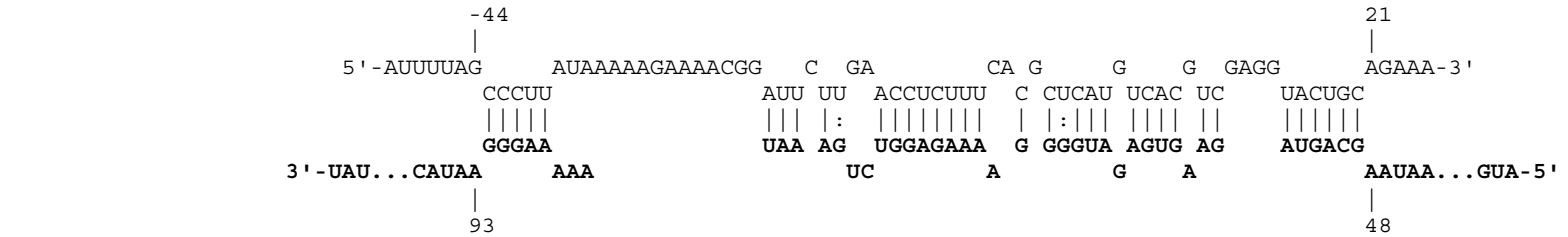
---

6. Predicted interaction between Hv_nc13 (T*rpoB*-2586) and *rps2*|Description:ribosomal_protein_S2|GeneID:4525123



Energy:	-15.3 kcal/mol	Position - mRNA:	-29 -- 5
Hybridization Energy:	-21.1 kcal/mol	Position - ncRNA:	23 -- 53
Unfolding Energy - mRNA:	2.9 kcal/mol	Position Seed - mRNA:	-16 -- -10
Unfolding Energy - ncRNA:	3.0 kcal/mol	Position Seed - ncRNA:	33 -- 39

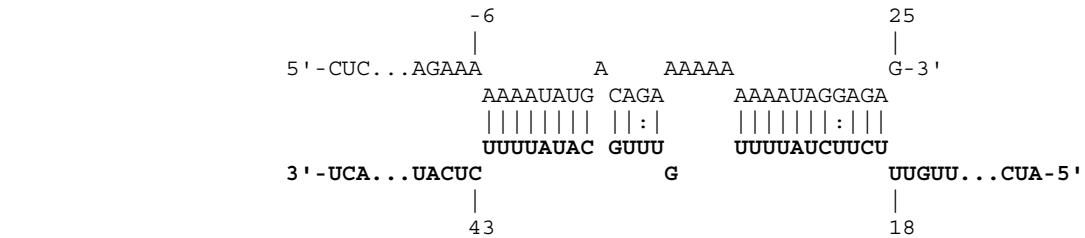
7. Predicted interaction between Hv_nc12 and *rps*7|Description:ribosomal_protein_S7|GeneID:4525104



Energy:	-15.0 kcal/mol	Position - mRNA:	-43 -- 20
Hybridization Energy:	-34.6 kcal/mol	Position - ncRNA:	49 -- 92
Unfolding Energy - mRNA:	10.8 kcal/mol	Position Seed - mRNA:	-14 -- -8
Unfolding Energy - ncRNA:	8.8 kcal/mol	Position Seed - ncRNA:	70 -- 76

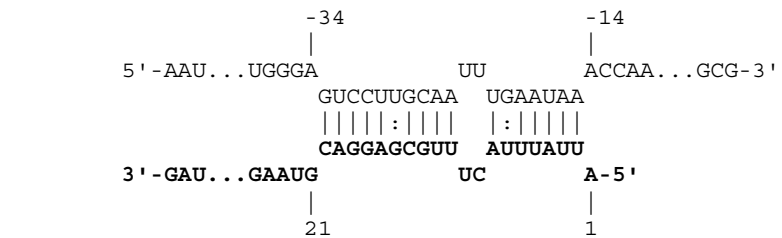
---

8. Predicted interaction between Hv_nc19 and *infA*|Description:translation_initiation_factor_1



Energy:	-14.7 kcal/mol	Position - mRNA:	-5 -- 24
Hybridization Energy:	-19.5 kcal/mol	Position - ncRNA:	19 -- 42
Unfolding Energy - mRNA:	0.7 kcal/mol	Position Seed - mRNA:	18 -- 24
Unfolding Energy - ncRNA:	4.1 kcal/mol	Position Seed - ncRNA:	19 -- 25

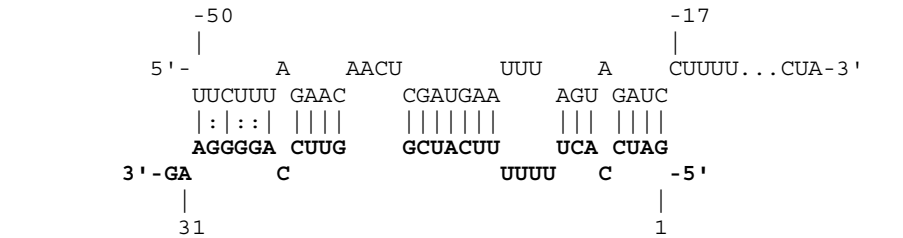
9. Predicted interaction between Hv_nc8 and *psbA*|Description:photosystem_II_protein_D1



Energy:	-14.5 kcal/mol	Position - mRNA:	-33 -- -15
Hybridization Energy:	-22.6 kcal/mol	Position - ncRNA:	2 -- 20
Unfolding Energy - mRNA:	2.7 kcal/mol	Position Seed - mRNA:	-21 -- -15
Unfolding Energy - ncRNA:	5.4 kcal/mol	Position Seed - ncRNA:	2 -- 8

---

10. Predicted interaction between Hv_nc49 and *psbF*|Description:photosystem_II_protein_VI



Energy:	-14.3 kcal/mol	Position - mRNA:	-50 -- -18
Hybridization Energy:	-21.5 kcal/mol	Position - ncRNA:	1 -- 30
Unfolding Energy - mRNA:	3.7 kcal/mol	Position Seed - mRNA:	-35 -- -29
Unfolding Energy - ncRNA:	3.4 kcal/mol	Position Seed - ncRNA:	13 -- 19

Appendix G: PSs revealed by dRNA-seq. PSs are named after the downstream located gene and the number of nucleotides between the 5'-P end and the start codon of the ORF. The genomic location and the strand of each PS, as well as the sequence 20 nt downstream are given. The abundance of PS is represented by the number of cDNAs in both green and white +/- libraries. Processed 5' ends which are represented by cDNAs that do not reach the downstream located gene are referred as "disconnected" in the *Comments* column. PSs mapped within 10 consecutive nucleotide were considered a single 5' end and denoted by the genomic position of the most abundant of the 5' ends in (-) library. The less abundant 5' ends are referred to as minor 5' ends and are listed in the *Comments* column. The references for the PSs which are in agreement with previously published data are given and the 5'-P ends experimentally verified in this study are marked in bold in the *Name* column. PSs, which do not represent 5' ends of downstream RNAs are underlined. PSs of rRNA precursors are highlighted. The rest of the PSs mark potential 5' processed termini of downstream mRNAs.

Name	Genomic position	Strand	Sequence PS + 20 nt downstream (21 nt)	No of cDNAs (G-/G+)	No of cDNAs (W-/W+)	Comments	Reference
PS*trnK*-449	4917	-	TATCGTGCCAATCCAACATAA	238/0	53/0	downstream of *rps16*	
PS*rps16* -79	6153	-	AAACCAATGACTATTCATGAT	29/0	12/0	minor 5' ends: 6153-6154	**5'-RACE**
PS*psbD*-135	9024	+	TTTCTCTCTTCGAGACCATTG	25/0	n.d.	minor 5' ends: 9023-9025; disconnected	Christopher, et al., 1992
PS*psbC*-45	10159	+	ATCAGCCTCATGAAAATCTTA	5/0	n.d.	GTG (10204) is used a start codon (Kuroda, et al., 2007)	
PS*rpoB*-126	19961	+	TAGAATTTCATGTGATTCAGT	7/0	4/0	disconnected	
PS*rps2*-97	30276	+	ATTTATTTCAAGCTATTTCGG	594/0	560/0		
PS*atpH*-49	32610	+	ATTGTATCCTTAACCATTTCT	2805/2	75/4		Pfalz, et al., 2009
PS*atpA*-1411	33425	+	TAATACCGATATTTTAGCAAC	53/0	1/0	minor 5' ends: 33423-33425; disconnected; within *atpF* ex. 1	
PS*atpA*-52	34784	+	ATTTAGGCATTATTTTTCCCC	5/0	1/0	minor 5' ends: 34783, 34784; disconnected	
PS*rps14*-59	37310	-	ATTTATTTTTCCATCTAGGAT	14/0	1/0		

PS*psaA*-584	42464	-	TTTCATTATATCCATTTCTTA	21/0	2/0	minor 5' ends: 42464- 42468; disconnected; downstream of *ycf3*	
PS*ycf3*-62	44608	-	TTTGTTTTTATGTTATTTTGT	195/0	105/0		**5'-RACE**
PS*trnT*-2006	48537	-	AACTTTGTATCGCGCACATGA	94/1	0/0	minor 5' ends: 48538- 48532; disconnected; downstream of *ndhJ*	
PS*ndhK*-57	50163	-	TTTCGTGCTTATCTTAGTTGT	152/0	81/2		**5'-RACE**
PS*rbcL*-59	54880	+	CATCGAGTAGACCCTGTTATT	226/0	1/0	minor 5' ends: 54879, 54880	**5'-RACE;** Reinbothe, et al., 1993;
PS*petL*-66	64018	+	CTTAGGTAAATGCTTTACCAA	20/0	5/0	minor 5' ends: 64018, 64019	**5'-RACE**
PS*rpl33*-161	65633	+	ATTGTATTCTTTAATTATTTC	155/0	3/0	minor 5' ends: 65623	Pfalz, et al., 2009
PS*rps12*-52	68160	-	ATCAGGTTAAGATGGATCTAA	24/0	17/0		**5'-RACE**
PS*psbB*-63	69348	+	TTTTTCAATGCGATAAAATAA	16/1	n.d.		**5'-RACE;** Westhoff, 1985;
PS*psbH*-37	71455	+	AGTATACAAAGTCAACACCAA	223/0	1/0		Felder, et al., 2001
PS*petB* -44	71799	+	GGTAGTTCGACCGCGGAATTT	20/0	6/0	disconnected	Pfalz, et al., 2009
PS*petD* -148	73282	+	ATATCGGGTAGGTTGTGGTAT	833/4	98/5	minor 5' ends: 73281, 73282	Barkan, et al., 1994
PS*trnH*-6834	74761	+	TTTTATTATGATCCATTTCGC	612/1	37/0	minor 5' ends: 74757-74762; disconnected; downstream of *petD*	Chen and Stern, 1991
PS*ndhB*-6	89040	-	TAATTCATGATCTGGCATGTA	284/0	23/0	minor 5' ends: 89050-89040	Hashimoto, et al., 2003
PS*ndhB*-173	89207	-	GAAATCATGATCAACTAAGCC	8/0	3/0	disconnected	**5'-RACE;** Hashimoto, et al., 2003
PS*rps12*-683	91341	-	CAACATAGGTCATCGAAAAGA	4/0	8/0	disconnected	
PS*rrn16*-28	92657	+	AAGGAAGCTATAAGTAATGCA	17/0	2/0		
PS*rrn23*-73	96389	+	TTCATGGACGTTGATAAGATC	191/0	60/0		
PS*psaC*-188	110094	-	CAAAATTCAAGTCTCTTGGCT	47/0	7/0		**cRT-PCR**
PS*ndhA*-67	114517	-	AAATTGGCTGATATCATGACG	26/0	4/0		María del Campo, et al., 2006
PS*ndhI*-49	112277	-	CCAAACAAGAGAAAGAAACAT	5/0	n.d.	disconected	

Appendix H: Processed mRNA termini in barley chloroplasts detected by dRNA-seq and verified by alternative methods. The PSs detected by dRNA-seq and the processed mRNA termini associated with them are listed. The genomic position of the 3' mRNA termini is deduced from the position of the 3' ends of the sRNAs associated with them. In most cases, sRNAs represent populations with ends mapping within several nucleotides. The numbers of clones supporting the genomic locations based on dRNA-seq, as well as the total number of sequenced clones (numbers after the slash) are given. The genomic location of additional ends detected by 5'/3'-RACE and the number of sequenced clones supporting them is presented.

Name	Processed Termini	Genomic position based on dRNA-seq	Strand	Alternative method	Clones supporting dRNA-seq data	Additional termini
PS*trnK*-449	*rps16* 3'	4890-4892	-	3'-RACE	4/5	
PS*rps16* -79	*rps16* 5'	6153-6154	-	5'-RACE	6/8	6152 (2/8)
PS*psaA*-584	*ycf3* 3'	42440-42444	-	3'-RACE	7/7	
PS*ycf3*-62	*rps4* 3'	44583-44588	-	3'-RACE	2/4	44574; 44577 (2/4)
PS*ycf3*-62	*ycf3* 5'	44608	-	5'-RACE	6/7	
PS*trnT*-2006	*ndhJ* 3'	48515-48517	-	3'-RACE	5/5	
PS*ndhK*-57	*ndhK* 5'	50163	-	5'-RACE	5/5	
PS*rbcL*-59	*rbcL* 5'	54880	+	5'-RACE	6/6	
PS*petL*-66	*petL* 5'	64018	+	5'-RACE	6/14	64020 (2/14)
PSrps12-52	*clpP* 3'	68131-68133	-	3'-RACE	7/7	
PS*rps12*-52	*rps12* 5'	68160	-	5'-RACE	6/8	
PS*psbB*-63	*psbB* 5'	69348	+	5'-RACE	12/15	
PS*trnH*-6834	*petD* 3'	74780-74783	+	3'-RACE	4/6	74748 (2/6)
PS*ndhB*-17	*rps7* 3' (long)	89023-89026	-	3'-RACE	7/7	
PS*ndhB*-173	*ndhB* 5' (long)	89207	-	5'-RACE	6/8	89206 (2/8)
PS*psaC*-188	*ndhE* 3'	110064-110069	-	3'-RACE	6/6	
PS*psaC*-188	*psaC* 5'	110094	-	cRT-PCR	5/6	
	psaC 3'	109563-109569	-	cRT-PCR	5/6	
	ndhD 3'	107917-107922	-	3'-RACE	4/6	

Appendix I: Potential mRNA 3' termini revealed by hairpin RNAs resistant to TEX treatment. TEX-resistant cDNA accumulations mapped near the 3' ends of 13 genes reveal potential mRNA 3' termini. These mRNA 3' ends are proposed to be generated *via* RNA structure-mediated blockage of 3' nucleases. The name and the genomic position of the end of the genes are given. The genomic positions of the most predominant 3' end of these cDNAs accumulations were selected as potential mRNA 3' ends and the corresponding 3' UTR lengths (nt) were calculated. The optimal secondary structure and the minimum free energy (kcal/mol) of the inverted repeat (IR)/stem-loop predicted near the potential mRNAs 3' ends were predicted using RNAfold Server (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Gene	Strand	Gene end	mRNA 3' end	3' UTR length	Potential inverted repeat (IR) near mRNA 3'end	Min. free energy (kcal/mol)	Comments/Reference
rps19	+	490	573	83	AAAAUACCCAAUAUCUUGCUAGAACAAGAUAUUGGGUAUUUU ((((((((((((((((((.....)))))))))))))))))	-23.1	
psbA	-	619	532	87	AAAAUACCCAAUAUCUUGUUCUAGCAAGAUAUUGGGUAUUUU ((((((((((((((((((....)))))))))))))))))	-25.2	Memon, et al., 1996
psbC	+	11589	11659	70	UGGCUCGGUUAUUCUAUCUAGCCGAGCCA ((((((((((.......))))))))))	-18	
psbM	+	17320	17456	136	UAAAGUGUGGUAGAAAGAACUACAUAUAGUUUUUUCUACGACACUUUA ((((((((.((((((((((((....))))))))))))).))))))))	-24.9	
rpoC1	+	25403	25461	58	UCGGCGAUGCCCCUCCCCUUUGCUUUCGGGGGGCAUUCCGA ((((.((((((((((.....)))))))))))))	-21.7	
rps14	-	36940	36822	118	CCCUCUUUACCAUUCUGUAUAAAUGGACUAUUCUAUUUGUAUAGAUAUGGUAGAGGG ((((((..((((((((((((((((((....))))))))))))))).))))))))))	-28.8	Kim, et al., 1993
rbcL	+	56378	56505	127	UCGGCUCAAUCUUUUUUUUUUAUAAAAAAAGAUUGAGCCGA ((((((((((((((((.....))))))))))))))))	-24.7	Calie and Manhart, 1994
petA	+	61333	61601	268	UCGGCACAAGAAAAAGGCUUUUUCUUGUGCCGA ((((((((((((((...))))))))))))))	-20.2	

psbJ	-	62154	62066	88	CGGGUCCUUACCCCCUUUAUCUGAUUAGAGCGGAAAGGACCCG ((((((((((..((.(((((.....)))))).)).))))))))))	-20.7	
rps18	+	66774	66897	123	UUCCCGGAGUUCCCUCUCCGGGAA ((((((((((.....))))))))))	-16.4	
psbT (psbN)	+	71208	71250	42	UAAGAAGUCUCCCAGAUAGGGGGACUUCUUA (((((((((((((.....)))))))))))))	-20.1	the stem loop structure maps downstream of *psbN* on the opposite strand; may stabilize the *psbN* mRNA as well
ndhD	-	108033	107918	115	UUGAGAACCCUUUGAGAAGGCGCUCAAGGGGUUCUCAA (((((((((((((((.....))))))))))))))))	-25.4	3'-RACE (Appendix H)
psaC	-	109622	109566	56	ACCGAAGAAGCCUGUGCUCGAAAUAAUCGAGCACGGGCUUUUCUGGU ((((.(((((((((((((((.....)))))))))))))))).))))	-31.5	cRT-PCR (Appendix H)

Appendix J: Potential mRNA 3' termini revealed by sRNAs resembling PPR/PPR-like protein footprints. Unstructured sRNAs resembling the footprints of PPR/PPR-like proteins mapped downstream of 14 genes reveal potential mRNA 3' termini. These mRNA 3' ends are proposed to be generated *via* protein-mediated blockage of 3' nucleases. The gene name, strand, and the gene end genomic position are listed. In most cases, sRNAs represent populations with ends mapping within several nucleotides. The genomic positions of the most predominant 3' end of the associated sRNAs were assigned as the potential mRNA 3' ends and the corresponding 3' UTR lengths (nt) were calculated. The references for the eight of the barley mRNA 3' ends, which are in agreement with 3' termini mapped in other species, are given. Six were experimentally verified as 3' termini in this study.

Gene	Strand	Gene end	mRNA 3' end	3' UTR length	sRNA associated mRNA 3' ends	Reference
rps16	-	5027	4890	137	TATCGTGCCAATCCAACATAAGCCCCT	3'-RACE (Appendix H)
atpI	+	32091	32632	541	ATTGTATCCTTAACCATTTCTTTT	Pfalz, et al., 2009; Prikryl, et al., 2011
atpF	+	34744	34805	61	AATTTAGGCATTATTTTTCCCCTT	Pfalz, et al., 2009
ycf3	-	42552	42443	109	AGAATTTCATTATATCCATTTCTTAT	3'-RACE (Appendix H)
rps4	-	45545	44585	960	TTTGTTTTTATGTTATTTTGTGAAG	Hattori and Sugita, 2009; Zhelyazkova, et al., 2011
ndhJ	-	48787	48515	272	AACTTTGTATCGCGCACATGACT	3'-RACE (Appendix H)
psaJ	+	65345	65656	311	ATTGTATTCTTTAATTATTTCTCT	Pfalz, et al., 2009; Prikryl, et al., 2011
clpP	-	68250	68126	124	ATCAGGTTAAGATGGATCTAAACCAATCCATTTTT	Meierhoff, et al., 2003; Zhelyazkova, et al., 2011
psbH	+	71713	71816	103	GGTAGTTCGACCGCGGAATT	Meierhoff, et al., 2003; Zhelyazkova, et al., 2011
petB	+	73243	73309	66	CATATCGGGTAGGTTGTGGTATTTCATTGCT	Barkan, et al., 1994; Fisk, et al., 1999
petD	+	74661	74780	119	ATTATTTTATTATGATCCATTTCGCG	3'-RACE (Appendix H)
rps7	-	89334	89023	311	ATGCAGTTACTAATTCATGATCTGGCATGT	3'-RACE (Appendix H)
ndhE	-	110387	110065	322	CAAAATTCAAGTCTCTTGGCTCTTTTCACGC	3'-RACE (Appendix H)
ndhA	-	112330	112255	75	CCCAAACAAGAGAAAGAAACATAT	Zhelyazkova, et al., 2011

7 ACKNOWLEDGEMENTS

I would like to especially thank Prof. Thomas Börner for giving me the possibility to do my doctoral studies in his lab group and work on such an exciting topic. I am grateful for his full support and guidance throughout my PhD, for our long discussions, which stimulated my scientific thinking, and for his professional and at the same time relaxed and friendly attitude.

I would like to thank the Deutsche Forschungsgemeinschaft (SFB 429) and the Helmholtz Graduate School "Molecular Cell Biology" at the Max Delbrück Center for Molecular Medicine for financing my PhD project, and the Humboldt University Berlin for providing the infrastructure for my work.

I would especially like to thank Prof. Jörg Vogel, Dr. Cynthia Sharma and Dr. Konrad Förstner for their invaluable support in deep sequencing and bioinformatics analysis.

I am very grateful to Prof. Alice Barkan for introducing me into the PPR world and for her important contribution to that part of my project.

I would also like to thank Dr. Karsten Liere for his willingness to discuss work-related issues at any time and his helpful comments.

I would like to thank all current and former colleagues at the Department of Genetics for creating such an enjoyable working atmosphere. Special thanks go to Dr. Yan Zubo for showing me many technical tricks, Conny Stock for keeping the lab "neat and tidy", and the lab3 girls Johanna, Lili, and Emilia for being always good for my mood.

I would also like to thank my parents Albena and Dimitar, my sister Maria and the rest of my loving family for their moral support and for their faith in me. I am grateful to my boyfriend Michael for encouraging me to follow a scientific career and giving me the feeling that nothing can ever go wrong. Last but not least, I want to express my gratitude to Freya the cocker for reminding me to also enjoy the simple things in life.

8 PUBLICATIONS AND CONFERENCE ABSTRACTS

PUBLICATIONS

Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J., and Borner, T. (2012). The Primary Transcriptome of Barley Chloroplasts: Numerous Noncoding RNAs and the Dominating Role of the Plastid-Encoded RNA Polymerase. Plant Cell doi: dx.doi.org/10.1105/tpc.111.089441.

Zhelyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suarez, M., Borner, T., and Barkan, A. (2011). Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. Nucleic Acids Res. doi: 10.1093/nar/gkr1137.

Berger, M., Farcas, A., Geertz, M., Zhelyazkova, P., Brix, K., Travers, A., and Muskhelishvili, G. (2009). Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. EMBO Rep. 11, 59-64.

Cabrera, R., Zhelyazkova, P., Galvis, L., and Fernandez-Lahore, M. (2008). Tailoring orthogonal proteomic routines to understand protein separation during ion exchange chromatography. J. Sep. Sci. 31, 2500-2510.

CONFERENCE PRESENTATIONS

Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J., and Borner, T. (2012). Small Genome and Complex Transcriptome: RNA-seq Analysis of Chloroplast Transcription. Plant and Animal Genome XX, San Diego, CA, USA.

Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J., and Borner, T. (2011). The primary transcriptome of chloroplasts revealed by dRNA-seq. Campus Buch Symposium, Berlin, Germany.

Zhelyazkova, P., Sharma, C.M., Liere, K., Vogel, J., and Borner, T. (2010). The primary transcriptome of chloroplasts determined by deep sequencing. 2nd International Symposium on Chloroplast Genomics and Genetic Engineering, Maynooth, Ireland.

CONFERENCE POSTERS

Zhelyazkova, P., Sharma, C.M., Liere, K., Vogel, J., and Borner, T. (2010). The primary transcriptome of chloroplasts revealed by deep sequencing. 12th MDC/FMP/HSR PhD Retreat, Rheinsberg, Germany.

Zhelyazkova, P., Sharma, C.M., Liere, K., Vogel, J., and Borner, T. (2010). The primary transcriptome of chloroplasts. 23rd Symposium on Plant Molecular Biology, Dabringhausen, Germany.

9 EIDESSTATTLICHE ERKLÄRUNG

Hiermit versichere ich, die vorliegende Dissertation eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben. Die dem Verfahren zugrunde liegende Promotionsordnung ist mir bekannt.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Berlin, 14 March 2012

……..

(Petya Zhelyazkova)